DOCUMENT RESUME

ED 319 425 JC 900 271

AUTHOR	McGee, Diane; Starnes, Christine
TITLE	Evaluation as Empowerment: Holistic Evaluation across the Curriculum.
INSTITUTION	John Abbott Coll., Sainte Anne de Bellevue (Quebec).
PUB DATE	Nov 88
NOTE	74p.
AVAILABLE FROM	Research and Development, John Abbott College, P.O. Box 2000, Ste. Anne de Bellevue, Quebec H9X 3L9 ($6.00).
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*College Faculty; Community Colleges; Content Area Writing; Evaluation Methods; Grading; *Holistic Evaluation; Questionnaires; School Surveys; Student Attitudes; Teacher Attitudes; Teacher Workshops; Two Year Colleges; Two Year College Students; *Writing Across the Curriculum; Writing Evaluation

ABSTRACT

A study was conducted at John Abbott College to establish the reliability and validity of holistic evaluation for assessing learning in content courses. The study involved: (1) an informal poll of members of 22 departments concerning the acceptability of holistic evaluation within their department; (2) a faculty survey to ascertain opinions concerning the issues of "writing-to-learn," "learning to write," and faculty collaboration on pedagogical matters; (3) a survey to determine student perceptions of the purpose and means of learning to write, processes of grading, and evaluation; (4) an examination of the impact of holistic evaluation on students' grades and teachers' attitudes toward the evaluation method; and (5) two faculty workshops on holistic evaluation attended by 22 teachers representing 11 disciplines. The grade analysis showed that grades determined through holistic and traditional methods were statistically no different. The faculty survey indicated that respondents in every division saw value in developing and encouraging writing, but were much less in agreement about what was wrong with student writing and the place of writing in the college curriculum. Students seemed confused about the purpose of writing and about writing and evaluation processes. After using holistic evaluation, most teachers were enthusiastic about the process and felt they would use it again. Sample holistic scoring guides and the survey instruments are appended. (AYC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319425

EVALUATION AS EMPOWERMENT: HOLISTIC EVALUATION ACROSS THE CURRICULUM

FINAL REPORT

NOVEMBER 1988

Diane McGee and Christine Starnes

JC900271

BEST COPY AVAILABLE

John Abbott College

Depot legal - premiere trimestre 1988
Bibliotheque Nationale du Quebec
Code de diffusion 1532-0229
ISBN 2-9800816-6-3

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. McGee

C. Starnes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvem
EDUCATIONAL RESOURCES INFORMAT
CENTER (ERIC)

☐ This document has been reproduced
received from the person or organiza
originating it.

☒ Minor changes have been made to imp
reproduction quality.

• Points of view or opinions stated in this d
ment do not necessarily represent of
OERI position or policy.

ACKNOWLEDGEMENTS

This research was funded by DGEC
through its programme for aid to
research on teaching and learning.

The study would not have been possible without the
committed co-operation of twenty-six colleagues from
thirteen departments of John Abbott College.

The project was strengthened by consultation with members of the Research Training Institute as well as by discussion within the Language Across the Curriculum Committee and the Department of English of John Abbott College. Statistical analysis was graciously provided by Hikmat Ajakie, Mathematics Department.

Additional copies of this report can be obtained by sending a cheque for six dollars to:

Research and Development

John Abbott College

PO 2000

Ste Anne de Bellevue

Quebec H9X 3L9

A summary of the report in English or French is also available for two dollars.

TABLE OF CONTENTS

Chapter I — Definition.....page 1

Chapter II — The Study.....page 7

Chapter III — The Findings - Quantitative Data....page 12

Chapter IV — The Findings - Ethnographic Data

A. Faculty Survey.....page 14

B. Student Questionnaire.....page 26

C. Receptiveness to Concepts.....page 32

D. Interviews.....page 34

Chapter V — Implications for Teaching.....page 43

Chapter VI — Diffusion of Findings.....page 52

Chapter VII — Budget.....page 54

Appendices

1. Publicity Flyer.....page 55
2. Sample Scoring Guides.....page 59
3. Faculty Survey.....page 62
4. Student Questionnaire.....page 65
5. Interview Questions.....page 67
6. Shared Grading Criteria.....page 68
7. Conversion Tables.....page 69

CHAPTER I: DEFINITION

This study dealt with a highly specific pedagogical technique, "holistic evaluation". As a result it is important to provide a full definition of this concept before either describing the project or reporting its findings.

Holistic Evaluation: Definition and Description

"Holistic evaluation" is a guided process in which a minimum of two readers respond immediately to a piece of writing and rank it, usually on a scale of 1 to 6, bearing in mind a set of criteria agreed upon earlier. Edward M. White opposes this "holisticism" to "analytic reductionism", which dissects the whole to analyse its parts but may miss the larger meaning. The reader relies on a sense of the paper as a whole, for, according to the Educational Testing Service, "the basic assumptions of holistic scoring are that each of the factors involved in writing skill is related to all the others and that no one factor can be separated from the others".

Careful selection of a topic is integral to the "holistic" process; James Hoetker documents a growing body of literature which points to a correlation between the content, form and wording of a topic or question and students' performance. In these terms,"holistic evaluation" can improve student performance by demanding that the teacher be explicit about the objective of the written exercise so that the essay question allows the student to give the kind of response desired.

Specifying the kind of response expected is the first step in defining criteria. Odell emphasizes that different tasks "may have to be judged by different criteria". These criteria, often called a rubric but in our project referred to as a scoring guide, must be established prior to rating the papers, possibly based on a pre-test of the topic.

"Holistic scoring" traditionally takes place during group marking sessions. A "controlled essay reading" (White) is crucial. At the beginning of the session, markers assess samples of papers in order to refine their rubric; therefore, they can refer at intervals to a collection of essays whose scores have been agreed upon.

Holistic Evaluation: Reliability

"Holistic evaluation" was introduced by the Educational Testing Service in Princeton, New Jersey based on the 1966 study by Godshalk et al. Since then it has been found to be

administrable on an extremely large scale such as for the English Composition Test at ETS, the New Jersey Basic Skills Assessment test, and the National Assessment of Educational Progress. It has not, until this project, been investigated in the classroom setting.

This process directly addresses the common objections to essay grading: subjectivity and lack of reliability. It gained prominence largely in reponse to Diederich's 1974 study showing the unreliability of traditional analytic assessment. White argues that "holistic evaluation" eliminates this unreliability in setting up an "interpretive community". Hence the importance of the "controlled essay reading", which both ensures that all exam readers are subject to the same constraints and distractions and allows them to work together with the same self-developed criteria. As long as this community's standards are confirmed and reconfirmed through group marking of sample papers and periodic reference to these anchor papers throughout the session, the reliability is assured. The fact that each paper is read twice and, in the event of substantial disagreement, reread strengthens this reliability. Reports on reliability vary from over 80% to 92% (Skzrec).

This reliability has been shown in Canada as well as in the U.S. "Holistic evaluation" has been used, as Evans shows, across Canada during the last decade; projects in Manitoba, British Columbia, New Brunswick and Ontario have refined its procedures. In Quebec, however, CADRE reported no published work on "holistic evaluation" even for mother tongue. We know informally, however, that is is used for an end-of semester essay exchange in the English Department of John Abbott and for placement purposes at Dawson College. It is, then, a well established process in Canadian English Departments.

Holistic Evaluation: Applicability

Our literature search shows that work has yet to be done applying this process to testing in disciplines other than mother tongue. In transferring this process to other disciplines we were investigating new ground. However, the advantages that have been shown for "holistic evaluation" of general writing competence strongly suggested that this process could be used effectively in other disciplines.

Its advantages for students are generalizable across disciplines. First, "holistic evaluaton" implies a certain kind of assignment. It is most successfully used to evaluate an essay that requires what Diane Braungart-Bloom calls "higher order thinking skills". The process reinforces the point that "writing is an exercise for the whole mind" (White), "not only as a medium of thought but also as a vehicle for developing it" (Braungart-Bloom, our underlining). Second, a student whose work is being evaluated holistically has a better sense of audience. If, as Michael Robertson claims, it is important for student writers to know that "a unique human reader" will be responding

to their work, then the students who know their work will be judged as a whole --not by a count of facts included or errors made-- will surely have a clearer sense of being engaged in a meaningful task.

Nor are the advantages for teachers in any way discipline specific. "Holistic" grading can be much more satisfying to the evaluator who is able to feel like what White calls a "live and sympathetic reader", not a computer. He also notes an increased confidence in markers as they see substantial agreement between themselves and other markers. In addition the group marking session establishes an "increased sense of community" as marking is transformed from an individual to a group responsibility. Skzrec interviewed teachers who called "holistic evaluation" fun and liked having "a support system". Conversely, "in isolation...assessment was a lonely, tedious job". The result is that participants become more able and willing to discuss openly their personal assumptions about evaluation and its implications for pedagogy. Undertaken and run in a spirit of collegiality, "holistic evaluation" is valuable professional development.

Pedagogical Assumptions

"Language is the exposed edge of thought."
-James Britton

The assumption underlying, even motivating, this study is that, because it is a powerful means of learning, writing should be integrated throughout the curriculum. Janet Emig's 1977 article most directly charts the parallels between the cognitive and affective demands of writing and the characteristics of effective learning as defined by leading thinkers such as Bruner, Vygotski, Luria, and Polyani. This article remains the best known articulation of the theoretic base of a growing number of Writing Across the Curriculum programmes, for which a formal network in the U.S.A. and Canada now exists. These programmes usually provide workshops and consultation for teachers of disciplines other than English (mother tongue), designed to facilitate the teaching of writing appropriate to their own subject. The Fulwiler and Walvoord texts are examples of many now available that describe Writing Across the Curriculum pedagogy, an approach which stresses writing as a process and the need for active learning.

The emphasis on active learning derives from the fact that the Writing/Learning Across the Curriculum rationale is not solely based on a coincidence of the characteristics of writing and effective learning. It is based on the fact, made known by the Bullock Commission after its extensive study in Great Britain, that "language has a unique rcle in developing human learning". Briefly, the argument is profoundly based on psychological and philosophical evidence of the interrelationship between language and thinking. Much of this has been integrated into James Britton's Language and Thought. We use language to

organize our perceptions of our experience, to make it meaningful. So when students write about history, for example, they are making meaning of --forcing themselves to understand-- the historical concepts that have been introduced.

Such richness of learning is often denied students when tests exclude essays. In the extreme example of multiple -choice testing, we see a format which, despite great sophistication of item writing and pre-testing, continues to focus not on making meaning but on recognition. In addition, this recognition is often triggered by clues. Bozo found that 44% of items had clues and that 70% of these could be correctly answered merely by applying those clues. Haladyna suggests that most "principles" of item writing are not empirically based. The 1984 two-page "Debate" article summarizes current findings: essays are used to measure more complex learning, particularly skills of analysis and synthesis. The counter-argument that multiple-choice tests allow students the maximum amount of time to think about, rather than produce, the answer draws a false distinction between thinking and writing and does not recognize how writing generates ideas and thinking. While various forms of short answer questions do require more recall, they too are heavily influenced by clues. Most importantly, by definition they focus on discrete facts and do not require the making of connections which both signals and creates understanding.

BIBLIOGRAPHY

_____. "The Great Multiple Choice Debate: Different Strokes for Different Folks", Innovation Abstracts, vol.VI, no.3 (February 3, 1984).

Atman, Kathryn S, Patrick Redick Romano. "Conation, Goal Achievement, Style and Holistic Education", American Educational Research Association paper, April, 1987.

Bamberg, Betty. "Multiple-Choice and Holistic Essay Scores: What Are They Measuring?", College Composition and Communication, 33 (December 1982), p. 404.

Bozo, William G. et al. "A Study of Test-Wiseness Clues in College and University Teacher-Made Tests with Inplications for Academic Assistance Centers". College Reading and Learning Assistance Technical Reports 84-01.

Braungart-Bloom, Diane. "Assessing Higher Thinking Skills Through Writing". Presentation at the annual meeting of the American Educational Research Association, April, 1986.

Britton, James. Language and Learning. Harmondsworth: Penguin, 1970.

Brown, Rexford. "Literacy of Thoughtfulness". Closing speech at National Testing Network of Writing Conference, Minneapolis, April, 1988.

Bruner, Jerome S. Toward a Theory of Instruction. Cambridge: Belknap Press of Harvard Univeristy, 1966.

Bullock Commission. A Language for Life. London: H.M.S.O., 1975.

Charney, David. "The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview", Research in the Teaching of English, vol. 18, no. 1, February, 1984.

Conlan, Gertrude. "How the Essay in the CEEB Test is Scored". ETS in-house document, 1985.

Cooper, Charles and Lee Odell. Evaluating Writing. Urbana, Illinois: N.C.T.E., 1977.

Emig, Janet. "Writing as a Mode of Learning", College Composition and Communication, 28 (May 1977), p. 122-28.

Evans, Peter J.A. "Canadian Activity in Large Scale Assessment of Writing", English Quarterly, vol. 18, no. 2 (Summer 1985), p. 23-37.

Freedman, Aviva and Ian Pringle. Writing in the College Years: Some Indices of Growth", College Composition and Communication, 31 (October 1980), p. 311-323.

Fulwiler, Toby and Art Young (eds.). Language Connections: Writing and Reading Across the Curriculum. Urbana, Illinois: N.C.T.E., 1982.

Gere, Anne Ruggles. "Written Composition: Toward a Theory of Evaluation", College English, vol.42, no.2 (September 1980).

Godschalk, Swineford, Coffman. The Measurement of Writing Ability. New York: CEEB, 1966.

Haladyna, Thomas M. and Steven M. Downing. "A Quantitative Review of Research on Multiple Choice Item Writing". Presentation at the annual meeting of American Educational Research Association, 1985.

Hettich, Paul. "The Evaluator Journal: A Qualitative Adjunct to Program Evaluation", CEDR Quarterly, vol. 13, no.2, Summer 1980.

Hoetker, James. "Essay Examination Topics and Students' Writing", College Composition and Communication, 33 (December 1982).

Hutchison, T.P. "Predicting Performance in Variants of the Multiple Choice Test". Presentation at the annual meeting of the Psychometiric Society and the Classification Societies, 1985.

Meyer, Paul. "Variables Affecting Holistic Evaluation". Presentation at College Composition and Communication Conference, Atlanta, March 1987.

Myers, Miles. A Procedure of Writing Assessment and Holistic Scoring. Urbana,Illinois: ERIC Clearinghouse on Reading and Communication and N.C.T.E., 1980.

Robertson, Michael. "Anybody Listening? Responding to Student Writing", College Composition and Communication, 37 (February 1986).

Scherer, Darlene. "Measuring the Measurements: A Study of the Evaluation of Writing". ERIC document, 1985.

Skrzec, Ann Elizabeth. "Writing Skills Assessment: A Campus Wide Approach". Presentation at the annual meeting of the Writing Centers Association East Central Conferences, May, 1985.

Vygotski, L. Thought and Language. Cambridge: M.I.T. Press, 1962.

Walvoord, Barbara E. Fassler. Helping Student Write Well: A Guide for Teachers in All Disciplines. New York: M.L.A., 1982.

White, Edward M. "Holisticism", College Composition and Communication, 35 (December, 1984).

- "How can what we know about testing help in the classroom?" Presentation at National Testing Network in Writing Conference, Minneapolis, April 1988.

Teaching and Assessing Writing. San Francisco: Jossey-Bass Publishers, 1985.

Witte, Stephen B. "Uneasiness about direct assessment". Presentation at National Testing Network in Writing Conference, Minneapolis, April 1988.

CHAPTER II: THE STUDY AND ITS OBJECTIVES

This action research project established the reliability and validity of holistic evaluation, defined in Chapter I, for evaluating learning in content courses. It looked at two aspects of holistic evaluation that had not previously been investigated. The literature documented the usefulness of the procedure for the assessment of general writing ability, not for evaluating specific learning of a discipline. And it described the highly controlled experience of mass testing, not the more natural classroom situation. In proving its hypothesis, the project also met its general objectives of encouraging the use of writing as a means of learning for students and of increasing the possiblility of collaborative pedagogical work between faculty.

These objectives derived from the definition of two problems.

First, because of the importance of writing as a means of learning (writing-to-learn) as described in Chapter I, it is essential that students be given written assignments in all courses. Yet often our current teaching and learning procedures do not reflect the inter-relationship between writing and thinking. We are not providing students with enough opportunities to use writing in order to discover meaning and thus to improve their learning. Often we do not ensure that either through its format or its feedback a test is a means of learning, not just of being evaluated. Nor are we consistently signifying the importance of writing by making a written component a significant factor in final evaluation.

Even in the considerable number of courses in our curriculum that do include a significant writing component, writing-to-learn does not ensue. In many cases this component is an end-of-semester term paper, in the composing of which the student is given little guidance. As a result, a distressing number of students produce poor writing, and teachers feel frustrated. Such problems are being addressed at John Abbott by the Language Across the Curriculum through its Bulletin and its workshops, which are making known alternative ways to assign writing. The holistic evaluation procedure reinforces such approaches. In other courses, however, there is no discursive writing, only the copying of notes or penning of short answers in tests of discrete facts. The making of connections that signals, indeed creates, understanding is not called for. Given course objectives that stress the transmission of a large corpus of factual information, teachers---even those committed to the principles of writing-to-learn---feel strongly that neither they nor their students have the time to incorporate any essay assignments into the course. In such cases, the introduction of holistic evaluation, either in

addition to or instead of multiple choice or short answer tests, can give students an added opportunity to test and deepen their understanding by writing.

The need for the type of active involvement in learning that writing demands has also been defined by our students. Le Vecu, la pratique et le concret dans l'enseignement collegial stressed the students' distrust and dislike of a passive absorption and regurgitation of facts. Writing demands a commitment that cannot be passive. In addition, one of the characteristics of the "concrete" learning that students are seeking is usefulness, and one useful subject they pointed to outside their areas of specialization or concentration was the study of writing. The fact that their definition of writing ability focused on grammar rather than expression of understanding reflects their lack of knowledge of current pedagogy and does not make their expressed need irrelevant to this argument.

The second problem that we were addressing is that many teachers feel isolated from their colleagues and would welcome the chance to work collaboratively on pedagogical matters. This is a major need in the network; the Gregoire Commission reported that only 35% of teachers are committed to their jobs while a staggering 52% are still searching for their identity. We interpret this to mean that they are seeking a meaningful role within their academic community. In a similar vein, the institutional analysis of John Abbott carried out in 1986 by the Research Training Institute pointed to a feeling of isolation and lack of recognition on the part of teachers. This suggests a strong need to establish a reason for, and mechanism of, working together to improve our role as instructors, the element of our professional lives from which the majority of us gain the most satisfaction. It is appropriate that such an interaction focus on evaluation, the area of the profession that these reports, as well as the current priorities of le Conseil des Colleges, define as most problematic for both teachers and students.

These problems on both sides of the teaching-learning interaction were addressed by the two elements of our study: research and intervention.

A. RESEARCH

The research element of the project first put the hypothesis in perspective by providing descriptive data and then evaluated it.

Descriptive

Our first form of research sought in three ways to define the "atmosphere" for writing in our college. This description furnished us with a sense of the attitude toward writing in all disciplines and helped us identify areas of concern which

teachers might later be interested in investigating.

1. Departmental Poll

Early in the Fall of 1987, we began the project by polling members of 22 departments, all of whom had previously shown interest in Language Across the Curriculum. The researchers informally discussed with these people the acceptability of holistic evaluation to their departments. We asked in particular about any constraints that might exist as a result either of Cahiers requirements or of departmental evaluation policies. These findings are reported in Section C of Chapter IV.

2. Faculty Survey (Appendix 3)

A questionnaire eliciting opinions on issues related to the project---writing-to-learn, learning to write, and faculty collaboration---was designed. Two sociologists were consulted on its design and it was pretested in two departments. It was circulated in the middle of the 87-3 semester to all faculty, of whom 158 (44%) responded. These replies were transcribed to allow a computer analysis upon which was based the interpretations in Section A of Chapter IV.

3. Student Questionnaire (Appendix 4)

A somewhat similar questionnaire was designed and pretested for the students. It focused only on issues directly relating to them: the purpose and means of learning to write, processes of grading and evaluation. In the winter semester of 1988 it was distributed only to students whose class was involved in this project. The use of a limited sample (430) was for two reasons. First it would have been logistically impossible to survey all John Abbott students. Second such a sample could provide more reflective answers. Since we knew that the students in these classes would be doing at least one written assignment, and that at least some of these teachers would be speaking to their students about the project, we could assume that these students would be thinking about writing. In fact, when we pretested the questions with a few students, we discovered that they took very seriously the task of participating in a research project, were most interested in giving thoughtful answers, and very much wanted to let teachers know how they felt about writing.

We wished participation in the project to disrupt teaching rhythms as little as possible. We did not, therefore, attempt to direct the distribution of the questionnaires. We know that most teachers had their students complete them in class. Also most teachers distributed them early in the semester, as we had encouraged, but distribution largely depended on when time was available in the course.

The analysis of this survey is in Section B of Chapter IV.

Evaluative

The second segment of our research, which took place at the end of the 87-88 academic year, evaluated the impact of holistic evaluation on the teachers and the students. The data derived from experiments by twenty-two volunteer teachers, preferably working in pairs, from eleven different departments who used holistic evaluation to grade a written assignment in one of their courses.

1. Grades

The grades for the particular holistic assignment, for the final evaluation of the course, and for a comparable previous course were collected and correlated, as reported in Chapter III.

2. Interviews (Appendix 5)

After each experiment with holistic evaluation, participants were interviewed by the two researchers. A structured set of questions was used to allow direct comparisons between the experiments, but there was also an open-ended component to ensure full expression of the participants' thoughts and feelings about this method of evaluation. These are fully reported in section D of Chapter IV.

B. INTERVENTION

This element of the study directly addressed the two general objectives of the project, that of fostering faculty collaboration and that of increasing the writing-to-learn opportunities of students. Not surprisingly, it grew to be the most vital element of the project, and as a result we gathered more ethnographic than quantitative data.

1. Recruitment

An action research project obviously depends upon participation. In some cases, the departmental poll immediately elicited volunteer participants. In others, the LAC contact person discussed the concept with colleagues. Where deemed appropriate (Sociology/Anthropology, Nursing, Dental Hygiene, Humanities, French, Psychology, History/Economics/Political Science), the researchers made presentations to department assemblies. The publicity in all cases centred on a flyer summarizing the concept and providing a sample question and scoring guide (Appendix 1) and on a ten-page article about the procedure (Edward White's "Holisticism").

As a result of this recruitment, 26 teachers representing 13 disciplines originally agreed to participate. 22 teachers, representing 11 disciplines completed the project.

2. Workshops

Participants were first aided in their experiment with holistic evaluation by two workshops. The first, held in December, dealt with designing written assignments for tests. It provided guidance in identifying the learning objectives to be tested and in designing questions that would directly evaluate those skills and/or that knowledge. Unfortunately an unavoidable postponement of the workshop made it impossible for the question designed at this workshop to be pretested with a first semster class. However, the general objectives of discussing writing to learn and of providing faculty interaction were most definitely fulfilled.

The second workshop focussed on grading criteria. While it provided less faculty interaction due to a crippling snowstorm, group meetings over the next two weeks compensated for this. The workshop itself met its objective of making explicit the instinctive criteria we all use for grading essays. In addition we were able to define certain criteria shared across disciplines. These (Appendix 6) were the base of discussion for the later group meetings.

3. Consultation

Follow up or make-up for the workshops was provided by the researchers whenever requested by the participants. In fact, we met with teachers from 9 disciplines to discuss revising either the question or the scoring guide.

4. Follow-up

As reported in Chapter VI, this will continue through distribution efforts during the 88-89 academic year, with distribution of this report, presentations, a follow-up meeting for participants, and Language Across the Curriculum activities.

CHAPTER III: QUANTITATIVE FINDINGS

COMPARISON OF GRADES

There was concern that the introduction of a new form of evaluation, even if only for a small component of the final grade, would adversely affect the students' marks. The table below records the grade averages for the courses in which the experiment took place, the Holistic Course, and those of a course with similar content and methods taught previously, in most cases in the winter semester of 1987. The value of the assignments within the final grade ranged from 1% (Team A) to 30% (Team E).As shown by the following analysis, the introduction of holistic evaluation did not lead to any differences of statistical significance.

	88-1 Holistic Course	Comparable Previous Course
Chemistry (A)	63	64
Mathematics (B)	68	60
Chemistry (C)	54	48
French (D)	75	--
	73	75
Sociology (E)	70	75
	68	--
	74	73
	69	69
	70	68
Nursing (F)	75	72
Physical Education (G)	77	71
	75	82
Humanities/English (H)	76	71
	70	71
	68	70
Anthropology/Humanities (I)	63	67
	61	70
	71	74
Fine Arts (J)	74	82*
Documentation Technology (K)	76	84*
English (L)	71	70
	74	--
	68	70
	68	70

*In both J's and K's class in 88-1 there was a heterogeneous group rather than a homogeneous group who had taken a similar course before; thus the teachers had predicted a lower class average.

Statistical Analysis

HYPOTHESIS 1: The mean of the course using the holistic method is the same as that ofthe previous course with traditional grading.
($H_o: U_1 - U_2 = 0$)

versus

HYPOTHESIS 2: Using the holistic method underestimates the students' performance hence reflecting a lower mean than would have been granted traditionally.
($H_a: U_1 - U_2 > 0$)

The test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{sp\sqrt{\frac{1}{n} + \frac{1}{n}}}$$

$$sp = \sqrt{\frac{(n-1)\,S_1^2 + (n-1)\,S_2^2}{(n_1 + n_2 - 2)}}$$

$$sp = \sqrt{\frac{(22-1)\ (56.59) + (25-1)\ (29.62)}{(25 + 22) - 2}} = 42.206 = 6.47$$

$$t = \frac{70.72 - 70.04}{6.47\sqrt{\frac{1}{22} + \frac{1}{25}}} = \frac{0.68}{1.89} = 0.35$$

Decision

We fail to reject the hypothesis of equality of the means of the grading methodologies being tested at the 95% level of probability. We conclude that the methods are numerically no different in assessing the worth of a student.

COMPARISON OF OTHER QUANTITATIVE MEASURES

The concern for grades having been allayed, there was no other numerical analysis appropriate for as naturalistic a study as this one. Although time logs were collected for both holistic and traditional grading in the hopes of doing an hourly "costing" of the systems, the participants were not at all confident about their accuracy, invalidating the logs as statistical data.

As the following chapters show, the richness of the qualitative findings more than compensates for the brevity of this chapter.

CHAPTER IV: ETHNOGRAPHIC DATA

A. FACULTY SURVEY

INTRODUCTION

This survey was sent to all faculty in the fall of 1987. The rate of response was very good: 154 completed questionnaires, for a total of 44% of the faculty on campus last fall. Of these, 36 people identified themselves as having previously taken part in Language Across the Curriculum activities. The divisional breakdown was as follows:

	Number of Responses	% of Total (154) Responses	% of Teachers per Division
Arts and Letters	35	23	67
Social Science	44	29	49
Science	39	25	33
Arts and Letters Technologies	5	3	15
Social Science Technologies	6	4	43
Science Technologies	25	16	47
Total	154	100%	44%

The purpose of the survey was to enable us to describe the "atmosphere" for writing in the college. This allowed us to have a sense of the attitude toward writing in all disciplines before beginning the project. We also hoped to identify areas of concern which teachers might later be interested in investigating. To this end, we grouped the questions asked into 6 categories: 1) Purpose of Writing, 2) Improving Student Writing, 3) Constraints on Teachers, 4) Purposes and Methods of Evaluation, 5) Grading Written Assignments, 6) Collaboration with Colleagues.

ANALYSIS OF RESULTS

PURPOSE OF WRITING (Questions 1, 4, 5, 9)

Faculty across the disciplines see writing as important to all students, not just those who will use writing in their future jobs. In particular, there is a widespread acceptance of the concept of writing as a learning tool as well as a means of classroom evaluation. A majority of respondants in all disciplines indicated that writing improves a student's understanding and that the main purpose of writing is not merely to show the teacher what one knows; this response was highest in Arts and Letters and Science Technologies.

1. Only students who will use writing in their future careers need to develop writing ability.

Agree: 1.3%, Disagree: 98.1%, Uncertain: .6%

The implication here is obvious: writing is viewed as important for non-vocational reasons. The rest of the responses in this category elucidate these reasons.

4. Students deepen their understanding of a topic when they write about it.

Agree: 88.6%, Disagree: 2.5%, Uncertain: 4.4%

Social Science Technologies had the lowest number in agreement with 66.7%.

9. After writing, one knows more about a topic than when one began.

Agree: 80.4%, Disagree: 7% Uncertain: 10.8%.

Arts and Letters, Social Science, and Arts and Letters Technologies all agreed at 80+%; the lowest agreement came from Social Science Technologies with 66.7%

As expected, the responses to this statement correlate with those to #4. Perhaps the difference of 8% in the rate of agreement can be accounted for by a distinction between knowing more and deepening understanding. There was also more indecision on this question.

5. The main purpose of having students write is to let them show the teacher what they know.

Agree: 13.9%, Disagree: 69%, Uncertain: 13.9%.

Arts & Letters teachers, at 74.3%, had one of the highest rates of disagreement with the statement, perhaps because writing assignments given in these departments tend to be, in James Kinneavy's terms, exploratory rather than informative. However, Science Technologies disagreed the most strongly at 84%. There is also a strong component of indecision, ranging up to 20% in Science and Arts and Letters Technologies.

To be consistent, there should be an inverse relationship between these responses and those to #4 and #9. That is, if students learn by writing, the main purpose of writing is not merely testing per se. Obviously, since our respondants are teachers, there is always a component of testing what the student's knowledge in our assignments; perhaps the indecision comes from the difficult issue of defining the main purpose of a writing assignment.

Conclusions

Respondants in every division see value in developing and encouraging writing. There is a particular interest in the intellectual skills that writing can help us to develop in our students. Teachers who have participated in Language Across the Curriculum workshops indicate this interest even more strongly.

IMPROVING STUDENT WRITING (Questions 2, 3, 7, 8)

The large majority of teachers see the teaching of writing as important and useful; however they are not necessarily certain of how one goes about teaching writing. Sixty percent of respondants feel that student writing would improve if students were given more chance to practice writing; Social Science and Social Science Technologies placed the least emphasis on practice. There was is some disagreement about the importance of grammar in the process of writing and of learning to write. There were very mixed responses to the questions of whether grammar is the major problem in poor student writing and whether a better knowledge of grammar will improve student writing. Although the majority of teachers feel that teaching grammar does not necessarily improve writing, this varies greatly from department to department, with Science Technologies and Arts and Letters least concerned about grammar. A number of people commented that grammar and spelling are big problems but that they weren't sure if they were the biggest problems.

2. If students are taught what one does when one writes, they can improve their writing.

Agree: 79.1%, Disagree: 1.9%, Uncertain: 11.4%.

The answers ranged from a low of 60% agreement in Arts & Letters Technologies to highs of 83.3% and 85.7% for Social Science Technolologies and Arts and Letters.

3. Grammar and spelling are the biggest problems in student writing.

Agree: 29.1%, Disagree: 44.3%, Uncertain: 23.4%.

The overall totals mean very little in this case, as this question gives us one of the widest differences across the disciplines:

Arts and Letters: 74.3% disagree
Social Science Technologies: 66.7% disagree
Social Sciences: 54.5% disagree
Science Technologies: 36% agree, 32% disagree, 24% uncertain
Science: 51.3% agree, 30.8% uncertain
Arts & Letters Technologies: 60% agree, 40% uncertain

It is clear that there is a wide divergence in views on the relative importance of grammar and spelling as compared to more global problems in student writing. The focus on mechanical

problems is particularly true in Sciences and Technologies. This response suggests the importance of LAC doing some educational work to alert faculty to other ways of assessing the problems in student writing.

8. Learning more grammar does not necessarily improve a student's writing.

Agree: 63.3%, Disagree: 22.2%, uncertain: 13.3%.

As with statement #3, Arts and Letters teachers were least concerned with grammar as a basis of writing: 77.1% agreed with the statement. Presumably this result is heavily influenced by the English Department where people might be more aware of recent pedagogical theories about the process approach to teaching writing. However, Science Technologies also agreed with the statement at 72%.

Other divisions were less definitely in agreement with the statement, thus indicating an assumption that more grammar study will improve writing: Sciences agreed at only 46.2% (35.9% disagreeing, 17.9% unsure), and Arts & Letters Technologies was equally divided at 40% between agreement and disagreement. In Social Science Technologies, 50% agreed and 33.3% were unsure.

7. Students can't write because they have had little practice in doing so.

Agree: 60.8%, Disagree: 12.7%, Uncertain: 24.1%.

We should note that this question could have been interpreted in two different ways: 1) that it is practice that makes a good writer, and 2) that high schools have not been giving students enough writing to do.

Again, the results varied from division to division, with Arts and Letters (71.4%) and Arts & Letters Technologies (80%) teachers agreeing the most strongly. Social Science and Social Science Technologies had the lowest rate of agreement, with 54.5% and 33.3% respectively. The strong elements of uncertainty presumably mean that teachers feel there are reasons other than lack of practice for students not being able to write.

Conclusions

The results of this group of questions are much more varied. If teachers are fairly clear and pretty much in agreement about the purpose of writing and the place of writing in the college curriculum, they are much less in agreement about what exactly is wrong with student writing and how it should be improved. Obviously, since most of the teachers responding do not teach writing per se, the amount of indecision and variation in this section reflects their unsureness about how to diagnose problems and teach in an area which is unfamiliar to them. Because teachers do value writing, there would seem to be some good possibilities for Language Across the Curriculum to intervene

here. Based on their responses, teachers who have participated in LAC seem to have a clearer sense of the writing process.

CONSTRAINTS ON TEACHERS (Questions 6, 10, 20, 25)

The majority of teachers do not feel constrained by the Cahier. Most teachers outside the English Department feel that they have enough time to help students with their writing; however a sizable minority, particularly in Science, Social Science and Social Science Technologies (25-33%) feel that they do not. Fifty-one per cent of teachers would assign more writing if they had more time. Although most teachers (69%) are confident about assessing students, there is again an uncertainty rate of 22%.

6. Teachers outside English departments do not have time to help students with their writing.

Agree: 19.6%, Disagree: 62%, Uncertain 15.2%

Disagreement ranged from 50 % in Social Science Tech to 80% in Arts and Letters Tech. The strongest agreement was 25% in both Science and Social Science and 33.3% in Social Science Tech. Thus, although the tendency was to indicate that all teachers do have time to help students with writing skills, it was only by a slight majority in Social Science, Science, and Social Science Technologies, where disagreement hovered around 50%. There is obviously a significant minority of teachers who feel that they do not have time. Overall, 20% agreed with the statement, but this result is skewed to some extent by the low 8.6% Arts and Letters agreement. The "uncertain" responses were also fairly large in this item, ranging from a low of 12% to a high of 20%.

10. I would like to assign more writing in my class, but I don't have time to grade it.

Agree: 51.3%, Disagree: 31.6%, Uncertain: 11.4%

Social Science, Science and Science Technology responses were consistent with their answers to #6 , with 56.8%, 56.4% and 36% respectively agreeing with the statement. 57.1% of Arts and Letters teachers also agreed that they do not have time to grade. However, Technologies teachers do not seem to feel that lack of time is what stops them from assigning more writing. Their agreement rates ranged from only 16.7% (Social Science Tech) to 36% (Science Tech).

In general, then, although teachers do feel time constraints, their agreement with the statement is not as emphatic as we might have expected. It is also possible that the high rate of uncertainty (11.4%) and/or disagreement (31.6%) could be interpreted as a statement that teachers are satisfied with the amount of writing they are presently assigning in their classes.

20. I feel confident assessing how well a student has fulfilled the objectives of my course.

Agree: 69%, Disagree: 7.6%, Uncertain: 22.2%

The degree of uncertainty in the responses to this statement was fairly high, ranging up to 32% in Science Technologies and 40% in Arts and Letters Technologies. Arts and Letters and Arts and Letters Tech had the lowest agreement rate at 60%. The low agreement in these divisions and the degree of uncertainty in others suggest that teachers might be open to discussions about evaluation and to experimentation with new forms of assessment.

25. Regulations in the Cahiers limit my ability to introduce innovative teaching.

Agree: 10.1%, Disagree: 71.5%, Uncertain: 16.5%

Responses varied from 60% disagreement in Arts and Letters Tech. to 83.3% in Social Science Related Technologies. Agreement was highest in Social Science at 13.6%.

Conclusions

The majority of teachers in every division did not respond that they do not have time to teach writing. As well, we see an openness to discussion and change that may allow us to encourage and help teachers to insert a writing component into their courses. The number of "uncertain" responses in this section suggests an interest in discussing issues involved with writing and assessment.

PURPOSES AND METHODS OF EVALUATION (Questions 11, 16, 17, 18, 19)

Teachers are not sure about the attitude of students toward objective versus written assignments but may tend to think that more studying is done for courses in disciplines other than their own. They feel mixed about the value of multiple choice tests, but strongly endorse written assignments as a method of evaluation. Teachers are quite definite that they want to evaluate the ability to synthesize, not just recall, facts.

11. Multiple-choice questions test understanding of the subject.

Agree: 33.5%, Disagree: 43%, Uncertain: 18.4

There was a very mixed response to this statement. Technologies tended to agree with it -- 48% in Science Technologies and 66.7% in Social Science Technologies -- except for Arts and Letters Technologies which had the lowest agreement rate of all at 20%. Thus, the 33.5% agreement rate overall can be broken down to 49% in technologies and only 29.6% in the academic divisions, with Science the highest at 36%. Conversely, disagreement ranged from 46% in the academic divisions to 25.3% in the technologies.

19. Students study more before an objective test than before an essay test.

Agree: 17.7%, Disagree: 32.9%, Uncertain: 48.7%

The rate of uncertainty is significant in the responses to this question, ranging from 20% in Arts & Letters Technologies to 59% in Science. Agreement ranged from a low of 10.3% in Science to a high of 40% in Arts & Letters Technologies. With the exception of Arts & Letters and Arts and Letters Technologies; however, more respondants disagreed with the statement than agreed.

Perhaps teachers in these Arts divisions, who might be less likely in fact to give objective tests, assume that students are studying more for tests in other disciplines. The inverse could be true of those disciplines that frequently use objective testing.

17. Written assignments are a valid means of evaluating students' learning.

Agree: 86.1%, Disagree: 1.9%, Uncertain: 7.6%

This is a neutral statement: not that writing assignments are the best or worst methods, simply that they are valid. It is not surprising, then, that most teachers agree. The responses ranged from a low of 66.7% agreement in Social Science Technologies to 100% in Arts & Letters Technologies. What is interesting is that the response to #11 was quite different; there seems to be more confidence in the validity of written assignments than in the validity of multiple choice.

16. Students should learn to make connections between the separate facts they learn in a course.

Agree: 95.6%, Disagree: 0%, Uncertain: 3.8%

This statement is not explicitly related to writing. However, if learning discrete facts is the major aim, then there is little argument for including a writing component in evaluation. To us, the response to this statement gives a clear message that writing assignments are appropriate in every area of the college. In Arts and Letters and in Social Science Technologies, agreement was 100%.

18. It is more important to test recall of facts than the ability to see relationships between those facts.

Agree: 0%, Disagree: 91.8%, Uncertain: 7.6%

We found the expected inverse correlation with question 16: teachers overwhelmingly disagreed with this statement. In fact, no one agreed with this statement, although there was some divergence in Social Science Technologies where 33.3% of respondants were uncertain.

Conclusions

The responses in this category indicate quite strongly that some written assignments would be appropriate in any discipline because of the value that most respondants placed on synthesis of course material. Former Language Across the Curriculum participants had more reservations about multiple choice testing and were more strongly in favour of written assignments.

GRADING WRITTEN ASSIGNMENTS (Questions 12, 13, 14, 15)

Across divisional barriers, teachers agree very strongly with a positive approach to grading. However, there is a good deal of uncertainty. In particular, there are very mixed responses about whether all errors should be corrected, whether form and content should be graded separately, and whether the same criteria should be used for every written assignment.

12. All the errors in student writing should be corrected by the teacher.

Agree: 31%, Disagree: 53.2%, Uncertain: 13.3%

Arts and Letters disagreed most strongly at 74.3%. Arts & Letters Tech were most in agreement at 100%. Social Science Tech was split evenly among the three categories.

These mixed responses suggest a real uncertainty about the role of "corrections" and about the responsibility of the teacher in evaluating written work.

13. The "English" and the content of written assignments should be graded separately.

Agree: 23.4%; Disagree: 55.7%; Uncertain 18.4%

The rate of disagreement varied from 28% in Science Technologies to 80% in Arts and Letters Technologies. The fact that 20% or more in all disciplines except Arts and Letters and Social Sciences were uncertain reinforces our overall impression of uncertainty about evaluation. The only really strong feeling that content and "English" should be graded separately came from Science Technologies at 48% in agreement. Thus, overall, these responses indicate some pre-existing openness to the concept of holisticism.

15. It is good to use the same criteria for grading every written assignment.

Agree: 30.4%, Disagree: 44.9%, Uncertain: 22.2%

The responses are again mixed. Arts and Letters disagreed most strongly at 68.6% with only 8.6% agreeing. But other divisions seemed quite unsure, with both agreement and disagreement in the

30-40% range. Arts and Letters Technology was somewhat of an anomaly with 60% agreement/40% disagreement rates.

These responses in general suggest a need to discuss scoring guides quite carefully with most teachers. It would be worthwhile to work toward clarifying and criticizing the assumed criteria for good writing.

14. Comments about strengths foster improvement as much as pointing out errors.

Agree: 91.8%, Disagree: 1.3%, Uncertain: 3.2%

Strong agreement, ranging from 79.5% in Science to 100% in Science Related Technologies. No issue here!

Conclusions

The general sense of unsureness about evaluation suggests that teachers should be open to discussion of some of these problematic areas. A discussion of the writing process, particularly the connections between form and content and the relationship of evaluation to a particular writing assignment would also be fruitful.

COLLABORATION WITH COLLEAGUES (Questions 21, 22, 23, 24, 26, 27, 28)

There is a definite interest in and acceptance of the concepts of collaboration: a solid majority of teachers are not afraid that it would undermine their pedagogical autonomy, enjoy discussing teaching methods, feel that exchanges on teaching methods are productive and would like more such exchanges. However, teachers have varying experience of the existing degree of cooperation; nearly equal numbers of respondants agreed, disagreed and were uncertain about whether their department fosters co-operation; only half of all respondants feel that their department has common goals, and only 35% feel that interaction within their department usually deals with pedagogical matters. It seems that despite a theoretical openness to co-operation and collaboration, problems seem to exist at the department level and there is some pessimism about co-operation within certain divisions. In general, teachers from Science Technologies seem to have the most favourable impression of the degree of co-operation within their departments, Social Sciences the least favourable.

21. Working in pairs or groups might undermine each teacher's autonomy in the classroom.

Agree: 8.2%, Disagree: 72.2%, Uncertain: 19%

The majority of teachers in every discipline disagreed with this statement, although a significant number were uncertain. Dis-

agreement ranged from 60% in Arts & Letters Technology to 83.3% in Social Science Technologies. In two divisions (Arts & Letters; Social Science Technologies) no one agreed with the statement. However, Arts and Letters had a relatively high number of uncertain responses (23%).

22. Interaction between members of my department usually deals with pedagogical matters.

Agree: 34.8%, Disagree: 48.1%, Uncertain: 13.9%

Agreement ranged from 0 in Social Science Technologies to 52% in Science Technologies. The other divisions were more consistent with agreement rates between 29% and 40%. Disagreement was highest in Social Science Technologies (66.7%) and lowest in Science (33.3%)

Even within the same division, there seem to be very different perceptions as to how much discussion of pedagogy goes on. And there are certainly some extreme perceptions of pedagogical isolation.

23. I enjoy discussing teaching methods.

Agree: 86.7%, Disagree: 3.8%, Uncertain: 7.6%

Science had rather a high number uncertain (20.5%) and Arts and Letters the highest disagreement (11.4%), but there was a relatively narrow range of agreement, from 77% to 98%.

This response makes an interesting complement to the responses to statement 22. It suggests that, even if pedagogy is not being discussed much, teachers in general enjoy such discussions when they occur. The implication is that teachers would like to discuss pedagogy more often.

24. Working in my department, I feel I am a member of a team with common goals.

Agree: 50%; Disagree: 25.9%, Uncertain: 22.8%

There are such major differences among divisions that these totals do not tell us much. Science Tech. teachers feel very much part of a team at 84%; Arts & Letters agree at 54.3%, but 20% are undecided. For Social Science, on the other hand, only 29.5% agree and 22.7% are undecided; Science has an agreement rate of 46.2% with a high 30.8% undecided. Arts and Letters Technologies shows equal numbers of teachers (40%) agreeing and disagreeing, with 20% undecided. Social Science Technologies are equally split between agreement and uncertainty.

This response is one of the most varied along divisional lines. In particular, the strong agreement in Science Technologies is striking. In some of these departments, the team approach is perhaps integral to the pedagogical approach and to the curricu-

lum itself. In other divisions, most notably Social Science, there seems to be a fair degree of alienation from the department, and/or a much more individualistic approach to the job of teaching.

26. Exchanges about teaching methods are usually not productive.

Agree: 6.3%, Disagree: 77.2%, Uncertain: 15.2%

Respondants disagreed strongly with this statement, ranging from a high of 96% in Science Tech. to lows of 71.4% and 71.8% in Arts & Letters and Science. In Science Technology, there is again a strong sense that cooperation is high. In most divisions, however, there was some uncertainty as well, ranging up to 23% in Sciences.

This response is less definite than that to #23, where 86.7% agreed. Teachers seem to be making a distinction between "enjoyable" and "productive." However, the attitude toward pedagogical discussion is still quite positive, despite the response to Question 22 which indicates that such exchanges do not occur very often.

27. My department does much to foster co-operation of teachers in the development of pedagogy.

Agree: 36.1; Disagree: 35.4%, Uncertain: 26.6%

A very mixed response, quite evenly distributed among the three categories. Arts and Letters and Science Tech. agreed at about 60%. In other departments the largest percentage of respondants disagreed. To some extent the differences correlate with the different responses to question 24, but the implied criticism of department cooperation is even stronger in this set of responses. Social Science, for instance, agreed at only 13.6% here, as compared to 29.5% in question #24.

28. At John Abbott, there should be more opportunities for exchanges on teaching methods.

Agree: 73.4%, Disagree: 3.2%, Uncertain: 19%

Moderately strong agreement, ranging from percentiles in the 80s for all the technologies to a low of 61.5% in Sciences. The reason for this last response is not clear, since 72% of Science respondants, in line with the other academic disciplines, agreed that exchanges about teaching methods are productive (#26). However, in the responses to #23, Science teachers were somewhat less enthusiastic about collaboration than their colleagues in Social Science and Arts and Letters, and in responses to #21 they showed more concern about working with other teachers than did their counterparts in the other two academic divisions.

Conclusions

It seems that there are different responses to theoretical statements about collaboration than to more specific ones. When the statement refers to "my department" there is a more negative attitude toward the possibility or experience of collaboration than when a more global statement is made.

CONCLUSION

We were extremely pleased with the response to the survey and also with the results. In particular, we wish to highlight the following points:

1) In general, the response to the survey was very positive. We felt encouraged both by the number of people who took the time to respond and by the substance of their responses. There was no indication of hostility toward the idea of writing being used as an assessment tool or of non-English Department faclty teaching writing. On the contrary, most teachers seem genuinely interested.

2) We do see a more strongly positive response to writing and a clearer understanding of how the writing process can work from past Language Across the Curriculum participants. This suggests the value of positive intervention in developing teachers' interest and confidence in teaching writing.

3) Despite their interest, teachers have some real "how to" questions about the process of teaching writing. We feel that this indicates the potential importance and usefulness of our project as well as of future Language Across the Curriculum activities. We see teachers' questions and uncertainty about the role of writing in their classroom and their own role as teachers of writing as very positive: that is, indicating a curiosity and openness toward these new ideas.

4) Teachers clearly indicated their interest in collaboration with colleagues. The challenge remains in implementing collaboration, since at this point it seems to be minimal and/or ineffective, from the point of view of most teachers. Again, the interest in collaboration indicates the potential usefulness of our project and of other LAC activities.

5) Overall, Arts and Letters and Science Technologies seemed to have given the most thought to writing pedagogy. However, respondants from all disciplines showed an interest in thinking about student writing and in discussion with colleagues. Most of these people have not attended LAC workshops, but the potential is there to encourage their more active participation. The responses of Arts and Letters Technologies were often at variance with those of the rest of the faculty; however, we must remember that only a small percentage of teachers in this division responded.

B. STUDENT QUESTIONNAIRE

INTRODUCTION

In the winter semester 1988 this questionnaire was distributed to students involved in the holistic evaluation project.

Of the teachers participating in the project, only one pair returned no questionnaires (Chemistry Team A). In some cases, however, only one member of the team returned them. A total of 430 questionnaires was returned. The breakdown is as follows: Anthropology/Humanities: 67, Business: 8, Chemistry (Team C): 57, Computer Science: 34, Documentation Technology: 29, English: 28, English/Humanities: 29, French: 51, Fine Arts: 21, Mathematics: 18, Nursing: 45, Physical Education: 18, Sociology: 24.

Since any class might include both students majoring in the subject and those taking the class for other reasons, we do not pretend that the opinions recorded here in any way represent the prevailing opinions of students in that discipline. However, we see some patterns in the responses to suggest that students, whatever their area of concentration, were influenced by the particular class they were in at the time they filled out the survey. For instance, a Math teacher distributed questionnaires at the end of the semester when the students had just completed the test which included the essay component to be holistically graded. Thus these students might have been thinking specifically of this exam when they responded to the survey. We also noticed that the responses from the Business class -- of which there were only 8 -- were often quite anomalous. We have indicated the range of results for each question but in the case of Business have always looked for correlation with another class.

In analyzing the responses, we divided the questions somewhat differently from the way they appeared on the survey given to students: Purpose of Writing (Questions 1, 4, 7, 10, 20); Improvement of Writing (Questions 2, 3, 5, 6, 9); Multiple Choice vs. Essay (Questions 11, 17, 18, 12); Evaluation (Questions 13, 14, 15, 16, 19, 8).

ANALYSIS OF RESULTS

PURPOSE OF WRITING

Perhaps the most interesting finding in this section is the degree of uncertainty with which students responded. This hovered near the 20% mark for all of the questions. There was also a distinct lack of agreement with one of our most basic assumptions about writing: that one learns more about something when one writes about it. More students see writing as just putting the contents of the mind on paper than see writing as a learning process. The correlation between the agreement rates for ques-

tions 20 and 7 suggests a quite formulaic definition of learning as a process of storing material in the mind rather than as a creative activity.

There is a very odd discrepency between questions 1 and 20, both of which state essentially the same thing. Between responding to the first and last questions in the survey, some students seem to have changed their mind that writing an essay does show the teacher what one has learned. Only one group of students (Phys. Ed.) was consistent in their responses; two groups (Nursing, Math) showed a rise of 20% between #1 and #20. It is, in fact, somewhat surprising that agreement with statements #1 and #20 is not higher. One would assume that the evaluative aspect of essay-writing would be almost universally clear to students. Those students who do not feel that essays let them show the teacher what they've learned must really see essay-writing as a pointless exercise.

Another example of inconsistency in this section is the contrast between the Sociology students' satisfaction in finishing a piece of writing (#10) and their relatively negative responses to statements about the purpose of writing (#1, 20, 4).

1. Written (essay) assignments let me show the teacher what I have learned.

Agree: 59.8% Uncertain: 20.7% Disagree: 17.4%
Range of Agreement: 72.2% (Phys. Ed.) - 45.8% (Sociology), 37.5% (Business)

20. When I put my ideas into an essay, I am showing what I have learned.

Agree: 69.8% Uncertain: 17.4% Disagree: 9.8%
Range of Agreement: 80% (Nursing) - 50% (Business), 54.2% (Sociology)

4. I learn about a topic when I write about it.

Agree: 56.7% Uncertain: 16.5% Disagree: 25.6%
Range of Agreement: 71.4% (Fine Arts) - 45.1% (French)

7. Writing is the process of putting on paper what is stored in your mind.

Agree: 64.9% Uncertain: 17.9% Disagree: 16%
Range of Agreement: 75% (English) - 54.9% (French), 37.5% (Business)

10. I feel satisfied when I have finished a piece of writing.

Agree: 65,3% Uncertain: 20.9% Disagree: 12.6%
Range of Agreement: 83.3% (Sociology), 75.9% (Doc. Tech.) - 55.6% (Math), 37.5% (Business)

IMPROVEMENT OF WRITING

There was again a lot of uncertainty in this category, ranging up to 33% for the statement about organization versus sentence structure. More significant, perhaps, than the actual "uncertain" responses, were the frequent close splits between agreement and disagreement. This was particularly the case in the statements asking students to decide whether sentence level correctness or overall organization is most important. It is perhaps worth noting that students in the French class emphasized sentences, grammar and spelling more than students in other classes. Perhaps we can see a correlation between these responses and those to #4 in the first section, which indicates that a relatively high percentage of these same students do not see learning as an outcome or purpose of writing. In other words, the emphasis on sentence-level correctness correlates with a lack of emphasis on global meaning.

Students agreed by only a bare margin that more practice would help their writing; interestingly enough, those most in agreement were in a subject (Math) where few essays are required; those least in agreement, with the exception of Business, were in a subject (English or Humanities) where several essays are written. However, students were more encouraging about the effect of teaching on their writing (#6) and are reasonably optimistic that their writing will improve while they are at college. Students in a science class were least hopeful about this.

2. **I would be able to write better essays if I did it more often.**

Agree: 56% Uncertain: 22.3% Disagree: 20%
Range of Agreement: 72.2% (Math) - 12.5% (Business), 41.4% (English/Humanities)

3. **Grammar and spelling errors are the worst weaknesses in writing.**

Agree: 44.7% Uncertain: 16.7% Disagree: 37.4%
Range of Agreement: 62.4% (Business), 54.9% (French) - 25% (Sociology)

5. **Organising an essay clearly is more important than writing "good" sentences.**

Agree: 39.8% Uncertain: 33% Disagree: 26%
Range of Agreement: 62.5% (Business), 58.6% (English/Humanities) - 21.6% (French)

6. **When I was shown what to do to write a paper, my writing improved.**

Agree: 62.1% Uncertain: 22.3% Disagree: 14.4%
Range of Agreement: 79.3% (Doc. Tech.) - 54.9% (French), 53.7% (Anthro/Humanities)

9. **I do not think my writing will improve in college.**

Agree: 16% Uncertain: 18.6% Disagree: 64.4%
Range of Disagreement: 81% (Fine Arts) - 50.9 (Chemistry)

MULTIPLE CHOICE VS. ESSAY

There is little indication of strong feelings for or against essay or multiple choice exams. In fact, there were more uncertain and/or mixed responses in this section than in any other. For instance, there were several responses to #17 of over 50% uncertain (Business, Phys. Ed., Sociology).

Some interesting details: In response to #18, Math students, writing fewer essays, seem to find them more taxing than do students in an English class, a case, perhaps, of preferring the usual method of evaluation to the less well-known. However, the responses on the relative fairness of the two methods (#11) do not follow divisional lines. We were pleased to discover that there is no strong feeling that essays are unfair.

11. **I study more before a multiple choice or short answer test than before an essay test.**

Agree: 38.8% Uncertain: 19.3% Disagree: 40.7%
Range of Agreement: 48.3% (English/Humanities) - 11.1% (Math), 12.5% (Business), 20.8% (Sociology), 28.6% (Fine Arts)

17. **Multiple-choice tests are fairer than essay examinations.**

Agree: 20.5% Uncertain: 40.2% Disagree: 37%
Range of Agreement: 35.6% (Nursing) - 0 (Business), 11.9% (Anthro/Humanities)

18. **I have to remember more in order to write an essay test or examination than to write a multiple choice test.**

Agree: 40.7% Uncertain: 19.1% Disagree: 37.4%
Range of Agreement: 66.7% (Math) - 21.4% (English)

12. **At the end of a multiple choice or short answer test I feel I haven't made use of all I know.**

Agree: 40.5% Uncertain: 28.1% Disagree: 28.6%
Range of Agreement: 49.3% (Anthro/Humanities) - 11.1% (Phys. Ed.)

EVALUATION

It is certainly encouraging that students claim to go over the corrections their teachers make on their papers. It is also encouraging that there is some degree of openness toward "English" counting in the evaluation of assignments outside of the English Department and that students perceive non-English Department teachers as caring about the quality of writing. Students also say quite definitely that they like positive responses and somewhat over half prefer separate marks for form and content.

There is quite a lot of uncertainty about the statements in this section too, however. Particularly strong uncertain responses were those to #16 -- most groups were uncertain at 40%+ -- and to #19, where 50% of Phys. Ed. students were uncertain and some others were in the 30% range. In the responses to #8, Chemistry students were uncertain at 33%, Math at 27.8%. It is interesting that these Science students do not know whether their teachers care about writing quality.

13. I like it when a teacher gives separate marks for the content of a paper and the quality of writing.

Agree: 62.3% Uncertain: 21.4% Disagree: 13.7%
Range of Agreement: 76.1% (Anthro/Humanities) - 50% (Business, Phys. Ed.)

14. It is helpful when teachers point out what I have done well.

Agree: 91.9% Uncertain: 3.5% Disagree: 2.1%
Range of Agreement: 100% (Phys. Ed., English/Humanities) - 75% (Business), 87.5% (Sociology)

15. I rarely go over the corrections a teacher makes on my paper.

Agree: 11.6% Uncertain: 13.5% Disagree: 72.3%
Range of Disagreement: 88.9% (Nursing) - 62.1% (English/Humanities)

16. Teachers tend to evaluate essays on the basis of their opinion of the students, not of the paper.

Agree: 15.8% Uncertain: 46.7% Disagree: 34.2%
Range of Disagreement: 48.3% (English/Humanities) - 12.5% (Business), 15.8% (Chemistry)

19. It is unfair for grammar, spelling and punctuation to count for marks except on English assignments.

Agree: 30.7% Uncertain: 26.5% Disagree: 40%
Range of Disagreement: 55.6% (Nursing) - 21.1% (Chemistry)

8. Teachers outside English departments do not care about the quality of your writing.

Agree: 11.6% Uncertain: 16.7% Disagree: 70.2%
Range of Disagreement: 88.9% (Phys. Ed.), 90.5% (Fine Arts) - 45.6% (Chemistry), 55.6% (Math)

CONCLUSION

In general, students seem quite confused about writing. This comes across to some degree in every category analyzed: in their understanding of the purpose of writing, of the writing process, and of the evaluation process. It shows in the large number of "uncertains" recorded, in significant equal splits between agreement and disagreement, and in inconsistencies between the answers to different questions.

Students do seem to generally accept the idea of writing being important, even outside English classes, and do not necessarily prefer objective evaluation. But they are not exactly sure what the purpose of writing is and certainly not how it is evaluated or how it can be improved.

Students must be receiving mixed messages from teachers as to what is and isn't important about writing. It would seem that teachers in all disciplines need to improve their own understanding of the writing process and clarify their expectations of student writing in order to be able to give the students a better sense of the purpose and process of good writing.

C. RECEPTIVENESS TO THE CONCEPTS

IMPORTANCE OF WRITING-TO-LEARN

The poll of the departments confirmed the findings of the faculty and student surveys. There is a general concern for writing in our college at the same time as a need to clarify how best to foster the development of student writing. Given the unstructured natures of the exchanges in this poll, however, a number of more specific issues were highlighted.

The Language Across the Curriculum respondents generally shared the feeling that there are courses where writing is appropriate and others where it is not. In a very few instances, a lack of writing was blamed on colleagues' laziness but usually the need to transmit an enormous amount of information over the short fifteen week period of one semester was deemed the cause. Hence, for example, the possiblility of more essays in complementary than in concentration science courses. As a result of the existence of these two camps, the researchers were able only to preach to the "converted". Those who feel writing does not fit their course simply tune out. It is hoped that the findings of the two surveys as well as the record of the values of holistic evaluation for our participants will begin to cut down this immediate barrier in the "other camp".

Two specific objections to emphasizing written assignments were aired. One was a need to train students for multiple choice testing. For example, the professional accreditation examination for nurses is multiple choice; their teachers would feel delinquent in not providing their students experience of this format throughout their six semesters. More commonly shared over departments was the time problem for students as well as faculty. There was considerable concern that our students are already overburdened with seven or more courses and that to increase the writing in one would be to add to the overall accumulation. Teachers also live with severe time constraints that do not allow them to add activities to their weekly schedules. Colleagues who declined to participate in our project were concerned not so much about the time needed to read papers as the supplementary time required for the collaborative aspects of holistic evaluation and for reporting to us.

HOLISTIC EVALUATION

Although there was general (as it turned out well-founded) scepticism about the time-saving element of holistic evaluation, there was the sense that it was an interesting notion, allied to current practices. Making known grading criteria beforehand as with a scoring guide, for example, parallels the circulation of a lab manual two weeks prior to a physics lab. Holistic reading---seeking the global meaning---is already frequently used, as our

participants attested. For example, a chemistry teacher reported using a holistic-type method: reading the conclusion of lab reports first, and then evaluating the report in terms of its own criteria.

There were, of course, objections to the concept. Particularly in the sciences there was concern that a holistic reading would be an inaccurate reading which would not fulfil their objective of evaluating precise, detailed information. Team C's report supported this hypothesis, but not on the basis of a rigorous holistic procedure. The other interviews show that accuracy of reading depends more on the teacher's experience with reading essays than on the form of evaluation. The second objection to holistic evaluation questioned not so much the procedure as the pedagogical usefulness of essay rather than short answer questions. Particularly in technologies where the course objectives focus on retention of discrete facts rather than on the ability to syntheszie or analyze, this is a valid point showing an area where neither holistic evaluation nor discursive writing are appropriate.

FACULTY COLLABORATION

Despite the concern about the amount of time that would be involved in exchanging papers, none of the teachers with who we discussed holistic evaluation raised objections to the collaborative part of it. Although these discussions, like the survey, made us aware of certain schisms within departments, in all divisions we saw endorsement of the theorectical value of working with a chosen colleague. In practice, this turned out to be a major incentive for participation in our project.

Overall, we were struck by the openness and willingness to experiment of the majority of our colleagues.

D. INTERVIEWS

The vast majority of our participants found holistic evaluation a valuable experience not only for their students but also for themselves. What the teachers learned from the experiment depended as much on the form of holistic evaluation adopted as on their previous experience. Thus the interviews are grouped in four blocks: unholistic/semi-holistic (A,B and C), standard holistic (D through G), interdisciplinary (H and I), and involving students (J,K,L).

UNHOLISTIC/SEMI-HOLISTIC

Team A (Chemistry)

A's experiment was only partially holistic. Although she had a second reader from her department, she designed the question (an issue to be debated by historical figures) alone and made no explicit scoring guide. The three-week take home group assignment that substituted for an early semester lecture quiz was worth 1% of the final grade, which must have done much to convey to the students the teacher's feeling of its lack of importance. Nevertheless, the essay---unusual in her discipline---did fulfil two course objectives: underwriting group work and giving some importance to a topic she had little time to cover in lectures. Having had little experience reading essays, she found the grading threatening. Even more threatening was the large number of discrepancies between her own grades and those recorded verso in pencil by her second reader.

Thus nothing was gained by the teacher except anxiety. We posit, however, that had there been real collaboration on the evaluation, beginning with making the criteria explicit, the discrepancies would not have occurred and the essays would have been, as for B, easier to read.

Team B (Mathematics)

This too was not a truly holistic procedure although there was a second reader. B designed his own question (a word problem) and drew up a check list of 6 elements to be included in the answer rather than a descriptive scoring guide. There was only one discrepancy between him and the second reader, other than two caused by plagiarism. He felt these discrepancies resulted from diffferent interpretations of what the student meant, probably the result of his lack of experience reading essays. He did not seek to resolve these discrepancies but used his own grade because the writer was his student and he saw the team marking as merely experimental. Neither of this team found reading the essays any more time-consuming than grading a regular exam, but both were uncomfortable with not correcting spelling and grammar. Still, they felt they could not penalize students for these errors since the check list covered only the student's grasp of subject material. Despite these constraints B planned to give more written assignments the following semester, maybe in

a more developmental sequence. He feels that word problems are more valuable because they parallel the "real world" of work where one is presented with problematic situations to contend with rather than being provided with numerical formulae.

Here too more collaborative work on the question and a more carefully developed rubric would have been useful although neither marker felt there were any serious problems with the experiment.

Team C (Chemistry)

While this team's experience was negative in terms of holistic evaluation, it was quite positive as an experiment in Writing Across the Curriculum. Their holistic grading of a short essay question worth 5% of a unit test worth, in its entirety, 5% of the final grade was unsuccessful. To prepare, they agreed on an essay question and devised a grading scheme rather than a descriptive scoring guide. They exchanged papers and compared grades, many of which were dissimilar. They felt this was due to their having been "slipshod" because they knew there would be a second reader, rather than attending to every last detail as their discipline demands. They did not discuss these discrepancies but arbitrated for their own students. Thus they reread most papers to decide on their own mark, the other's, or an average. At the same time they signalled errors in order to make the test a learning experience.

Thus they confirmed their preconception that holistic evaluation was an inappropriate procedure for their discipline. It is interesting to note, however, that they continued to go against departmental norms and assign varied writing: 8 journals in one class and several essays in the other. They also received departmental approval to include an essay question in the final examination of this multi-sectioned course. This was marked on a 6-point scale, but with a check list rather than holistically. The teachers were gratified to find that students in their two classes did somewhat better on this question than those in classes that had only written lab reports. The table below shows the percentage of low (0) and high (6) scores achieved by Team C's students, who had written substantially throughout the semester, compared to those of students who had not done much writing.

Essay Grade	C's	Others'
6	2%	2%
5	18%	16%
4	20%	21%
3	27%	21%
2	16%	22%
1	11%	13%
0	5%	5%

Team D (French)

This experiment was seen by the two participants as demanding but not as a failure. The pair taught a similar course and were able to agree on choices of open-ended topics to be used for an in-class essay near the end of the semester. Students were allowed to prepare for these and in one class to bring a draft. The team adapted the English Department's general scoring guide to their purposes. The writers' own teacher was the first reader and, given the department's tradition of individualism, maintained the right to overrule any disagreements. This individualism as well as organizational problems exacerbated by illness made the collaborative aspects of the experiment difficult. Also, both teachers found the holistic reading itself---looking for meaning rather than structures as they had been used to---very demanding. Yet they clearly defined two values: confirming one's grades, particularly at those moments when one feels one is losing perspective, and clarifying the reasons for having students write essays. These values were deemed strong enough to suggest a developmental sequence of assignments the next time the course is taught.

Holistic evaluation was thus deemed not only appropriate for the course but also, as in all the following cases, a means of learning about teaching.

Team E (Sociology)

Like Team F, E had often worked together before but still found problems coordinating the activities in this experiment. Again like F, they revised a question (in this case a general concept to be analyzed) that had previously been used for a take-home research paper. They also adopted a scoring guide one of them had used before. This was discussed at the beginning of a meeting where they graded anchor papers. Their reading was the most rigorously blind one. Although they used an objective coding system, appropriate to their discipline, this did not mask whose class the papers came from, and they found they graded their own students more severely than the other's. Having agreed on the anchor papers, they worked independently, one of them feeling guilty about not correcting spelling. Later they compared grades by phone. The few discrepancies they found resulted more from the complexities of the papers having been group projects than from the holistic procedure. The scoring guide was given to the students when the papers, along with two sets of comments---all, following one of the premises of holistic evaluation, being positive rather than negative---were returned. Nevertheless, the teachers felt the scoring guide needed to be further revised so that it could be given to the students as an incentive to improve their writing before they wrote the essay The teachers also felt the procedure would have been more useful at the end of the semester.

Although Team E---both experienced teachers---said they felt the procedure would be more useful for new teachers who have not yet internalized their grading standards, they found value for themselves in using the scoring guide to communicate with their students. They also found, as did most of our participants, that the procedure had validated their own grading standards in a way that their usual exchanges, frequent as they are, did not.

Team F (Nursing)

F was a truly holistic team although only one teacher's class was involved. The second reader had, however, given some lectures to the class and collaborated with the teacher to set up the end-of-semester in-class essay. The students had four weeks to research four questions demanding specific information, and they knew they would have to choose one of two on the day of the test. Although the team had used similar questions before, they heavily revised them particularly after investing much time drawing up a scoring guide. Even after those efforts they realized when beginning to read papers that their guide was too analytic, too much a check list, and that it needed refinement in order to allow a global assessment. Following the standard model more closely than any others except Team I, they began by reading the papers together. Although this method did enable them to deal with discrepancies as they arose, they abandoned it as too time-consuming because they talked so much. They also felt that reading separately allowed them to gain momentum and thus to mark more quickly and effectively. Most discrepancies between their grades were easily resolved by reference to anchor papers. Usually one reader had been "off the wall", probably because she read some papers late in the evening.

This coordination was one of the problems they saw in the procedure; the other was that of translating the scores into percentages. Their 0-5 scale represented blocks of 20%, which made resolving pass/fail discrepancies "horrific" and did not allow them to make sufficiently fine distinctions among the papers. They partially resolved this problem by using half values. Despite these problems, however, they were satisfied by having developed much better criteria for an improved scoring guide.

Although they had often collaborated before, they had learned from this experiment a new, "less nit-picking" way of reading and a new way of specifying criteria. They felt this would be more useful than their usual check list but disagreed between themselves whether it would be appropriate only for final tests.

Team G (Physical Education)

G was truly holistic and involved both teachers' classes. The team worked closely together in setting up the assignment, which was not a test but a final synthesis evaluating an activity. This work involved at least three revisions of the

questions and the general scoring guide, discussions of which involved considerable reflection on and clarification of the course objectives. Fulfilling a long-term objective of the project, the final assignment was not given in isolation but was the last in a developmental sequence of journal entries. Students had a week to write their final entry. The teachers exchanged the papers for a blind reading, each reading the other teacher's papers first, a method which they felt led to greater objectivity. There were few discrepancies. The one they did have to resolve brought forth further clarification of objectives, one of the major values of the experiment that they articulated. Equally important to them was that the students had learned more and had focused on significant elements of the activity because of the assignment. That is, the writing had made the students more critical. The constraints they defined concerned problems relating to grading journals, not to this particular form of evaluation. They felt holistic evaluation itself had speeded their reading, after the first five or six papers, their reading and had helped them see both the topic and the students more globally.

They had made themselves more aware of their grading criteria. They had also validated those standards by seeing how close these were to their partner's. As a result, they felt they had learned enough from the experiment to use such a system of grading independently in the future. This emphasis on initiation echoes the conclusion of a member of Team E.

INTERDISCIPLINARY

Team H (English/Humanities)

H was an interdisciplinary team who, for a mid-term test, gave to their two classes different questions eliciting material that each of the pair was familiar with. Separately, they prepared their own question(s) and scoring guide. One team member had an open-ended question which allowed "creativity", the other a much more directed question. The latter allowed for a specific scoring guide whereas the open-ended question had a two-part rubric, a general guide followed by a check list of facts for the answers. Designing both forms of guides was very time consuming. The teachers met to explain their questions and guides to each other prior to independently reading the essays that had been written in class. One of the pair had given students four questions to prepare, with the knowledge that they would have to write on one; the other put short answer questions at the beginning of the test as a form of warm-up to the essay. For the first exchange of student papers, the students' teacher was the first reader. This proved undesirable because it decreased objectivity and made it impossible to write comments on the paper itself. Comments were felt necessary as this was a mid-term test, but both found that recording them undercut any time saved by a holistic reading. The papers were returned to one set of students accompanied by the scoring guide; the other group of students discussed a model answer.

Problems identified by this team focused on the time demands, including the fact that the exchange had necessitated a delayed return of the papers to the students. They also pointed to general problems concerning a scoring guide. It can be "pie in the sky" rather than based on a current student reality, and it has to be drawn up ahead of time which might lead to "teaching to the test".

Nevertheless, the guide had the values (for one) of predicting problems with the question (which he then revised) and (for the other) of forcing consideration of appropriate questions to ask. The greatest value for them both, however, was the interdisciplinary exchange. This was so strong that after the essay exchange, they read each other's course texts. The "holistic evaluation" itself was for one of the pair a learning experience that helped her focus on the argument students were proposing. She did, however, stress that this lesson could have been much more directly, less time-consumingly learned. For her partner the experiment was a validation of his own sense of grading. He also discovered that his usual method of grading is quite similar to the holistic method, although without the formal scoring guide.

Team I (Anthropology/Humanities)

This second interdisciplinary team was different from Team H in that the pair had often worked together. Also the nature of their courses made it possible for them to give very similar questions; these required the same sort of analysis but of different data. They adapted our project's specific scoring guide to their question but, like J, were uneasy with the relative weighting of data recall and analysis. The students wrote the essays in class after two weeks' preparation, only one of the classes being allowed to bring notes. Like team F, I followed the traditional model and graded the papers together. Each read two papers and gave them to the other before they compared grades for the set of four. Fortuitously, the first set of four included a couple of excellent papers which gave both teachers a sense of the top range of marks. While reading they, like E, kept notes in order to justify their decisions. Discrepancies arising from these decisions were resolved by reference to the scoring guide. One of the discrepancies presented an unusual problem. The paper's major weakness was of penmanship. This had impeded one of the team's reading to such an extent that he could not follow the reasoning that the other had "seen". The second reader agreed to lower his mark on the grounds that such an impediment was a negative element even if not mentioned in the scoring guide.

In retrospect Team I felt they had learned much. Unlike E, they felt the procedure would have been most useful at mid-term when students could be given the scoring guide as a nutshell diagnosis of their strengths and weaknesses. The problem of weighting the value of data recall versus analysis that had been highlighted for them by the process of creating and applying a

scoring guide suggested to them that students need training in analysis, not just in retrieving data. As accustomed as this pair was to working together, they found that the holistic evaluation procedure had reinforced them by "adjusting their personal scoring guides" at the same time as showing that they had not "lost perspective" and were not "asking too much". They even suggested the process could be a means of avoiding burn out.

INVOLVING STUDENTS

Team J (Fine Arts)

Like K, J works in a small department and teaches a unique course, so she made her students the second reader. She designed the question---a very open-ended one for which, as with one member of team H, there was no right or wrong answer. That fact undoubtedly explains some of the discomfort she felt with the scoring guide, which was specific for the top scale but generalized thereafter. She was particularly unsure about her mid-range marks. Her discomfort also reflected the data/analysis weighting dilemma found by team I.

Students had not seen the question prior to writing but had had its conceptual demands explained. They were also encouraged to spend the first five minutes of the test thinking, not writing. The teacher graded the papers on her 8-point scale and in the following class distributed copies to pairs of students, making sure writers did not read their own scripts. However, unlike the student reader segment in L's experiment, this second reading did not increase reliablility. The teacher's average on an 8-point scale was 4.7, the students' 6. The teacher felt this was because, knowing their own papers were in the pile, the students realized that if the one they were reading got a 4, for example, their own would get a 3, and so would over generously give a 5. The objectivity and the sense of range built into L's procedure were lacking. This trend could have been exacerbated by the "fuzziness" (generality) of the scoring guide although the students said they found it perfectly clear.

Despite all these drawbacks, the teacher felt the students had learned from seeing interpretations of the question other than their own and that it had been good for them to know that there are standardized criteria "across the board". Although there were elements in the procedure that she would revise, she had good feelings about it and felt she had really learned from the very difficult task of "putting instincts on paper" that had constituted designing the scoring guide. She regretted not having been able to collaborate on the project.

Team K (Documentation Technology)

Also being in a small department and giving a unique course, K had to design the question and scoring guide alone and then work with her students on the grading. She had pretested this process mid-semester, so found drawing up a general scoring guide

straightforward. She also confirmed the feeling of many participants that making a guide the second time would be considerably less time consuming. Still, she was not completely satisfied with her guide. She realized, though, that her discomfort stemmed from her lack of decision as to the importance of referring to theory in the answer.

After she had graded the full set of essays, pairs of students used the guide in order to grade two or three papers. They had been told beforehand that their grade could raise but not lower the teacher's assessment. The few discrepancies that occurred probably resulted from the students' not having read a range of papers. Unusually, 7 of her 22 students did not want a peer to see their paper, and the teacher respected their wishes. She felt that in future she would assign a less personal topic if peer evaluation were to be used.

Reflecting on the experiment suggested to her ways to improve both the mechanism for such an exchange and a way to develop student writing---providing models. The scoring guide had helped her overcome the normal "agony" of deciding on criteria, and knowing that there would be a second reader had speeded her own reading time. She and G were the only participants to find the process saved time, a more evident benefit for K who did not have to read a colleague's set of papers in exchange. Furthermore, she liked having had the opportunity to open the evaluation system to her students. In short, K saw the process of "holistic evaluation" as new and different and made use of it to overcome problems she had previously identified in her teaching. Thus, she planned to use the system in the following semester.

Team L (English)

Since each member of this team taught two sections of the same course, they were able to establish a particularly interesting procedure involving students. To prepare for the evaluation, the teachers agreed to a shared question (a book review to be written in class with each student critiquing a different text). They used the English Department's general scoring guide but agreed before grading on specific differences between a 3 (fail) and 4 (pass): a 3 either only retold the story or ignored the aesthetic component whereas a 4 was truly, if skimpily, a review.

The students were prepared for writing by discussion of the format of a review and some analysis of models. The teachers independently graded the papers, discussed together any discrepancies and resolved these by reference to anchor papers. The class teacher then took section 1's papers to section 2 (and vice versa) where she read, in random order the anchor papers for each score. After the criteria had been discussed, pairs of students read a paper and had to reach consensus on a grade for it. They also filled in comment cards (as had the teachers) before calling out the marks to the teacher. All participants were "staggered" by the high correlation between teacher and

student grades. The fact that the graders, unlike those of J and K, were in a different class from the writers helped objectivity, and the fact that students had a chance to internalize the criteria by reading a range of papers also increased reliability. The two teachers gave this assignment at different points in the semester. When they undertook the process for the second time, it was even smoother and quicker.

The team found both setting up the system and the double reading time consuming. Still, they felt it had helped clarify both the objectives of the test and the criteria for the question. In addition, like G, they found that a holistic reading was more objective than regular grading in that it focusses exclusively on the course material and does not involve diagnosis of the individual student's strengths and weaknesses. In short, both members of the team felt the experiment had validated their marking.

Their marking system was also validated in the students' eyes, yet another advantage. The students found their own mark more meaningful after having had to decide on someone else's. In particular, one or two students who initially said their lower-than-desired grade was the result of a difference of opinion with the teacher seemed truly convinced that they had to confront some writing weaknesses. All students appreciated receiving three sets of comments. Finally, learning had been enhanced because the students had been given the opportunity further to analyze the question.

CONCLUSION

Overall, despite the amount of time required for holistic evaluation, teachers were enthusiastic about the process and felt they would use it again, in one way or another. They felt that the time burden was due to their lack of familiarity with the procedures and that ultimately a holistic reading could save time. Most importantly, they felt strongly that the time demands were outweighed by the benefits accrued: enhanced teaching, improved student learning, and---above all---collaboration.

CHAPTER V: IMPLICATIONS FOR TEACHING

An examination of the application of holisitic grading to the classroom is the most significant part of this report. For, although participants found the process relatively time-consuming, they all felt positively about it and stated that the experiment had been worthwhile in a number of different ways. We shall, therefore, summarize both the values and limitations perceived by our participants and use these pros and cons as the base of three models for classroom use: one for two or more teachers, one for a single teacher, and one for teachers and students.

PERCEIVED VALUES

Participation in the project was almost universally seen as a valuable exercise. Even more importantly, nearly all the participants were enthusiastic about the holistic method itself and interested in using it again. The different values they perceived help focus on the implications of holistic evaluation for future teaching.

Benefits to Teachers

The strongest consensus among our participants was that the use of holistic evaluation enhances teaching. It gave them the opportunity to analyze and to articulate their own pedagogical values and practices. It deepened awareness of the processes of writing and learning, thereby making explicit some latent assumptions about pedagogy.

It is interesting to note that while holisticism mainly involves a new method of evaluation, it also demands a modification in the preparation for evaluation. In designing the scoring guide, teachers articulate their own assessment values and usually negotiate them with another teacher. This helps to clarify the objectives both of the course as a whole and of the test in particular. As a result of this clearer perception, the essay question itself can be revised, with positive results---as shown below---for student learning. In particular, teachers have to focus on different types of questions and on what can be tested by each. Such clarifications are vital to effective teaching. The work that has been done, for example by Hoetker and Freedman, defining the different cognitive demands made by different essay prompts has shown that we are often not serious enough in formulating questions. "Discuss X" can often test a myriad of abilities and pieces of knowledge that may not be linked to our course objectives. The meaningful revision of essay assignments and expectations more than compensates for the time that building a holistic evaluation scoring guide demands.

The re-evaluation of the assignment, expectations, and criteria in turn suggests means of improving methods of teaching prior to the testing. Providing models of strong and weak student answers, encouraging rewrites, constructing a developmental sequence of assignments were some approaches our participants suggested.

In addition, a holistic way of reading provides greater professional satisfaction. Not only is it "less picky" but also it is deemed to be a more meaningful process, enabling a more global view both of the topic tested and of the students themselves.

Perhaps the most important teaching gain is the confirmation of the teachers' own grading system. Whether comparing themselves with the scoring guide, another teacher, or the students, our participants found their judgements validated at the same time as their "instincts" were made more explicit. Given the amount of time that teachers devote to evaluation and the lack of certainty in this realm shown by our survey, the value of this bolstered confidence cannot be underestimated.

The single most valauble benefit of the project for most people was collaboration. Their enthusiastic response to working together echoed the interest in pedagogical cooperation expressed in responses to our survey. Designing the scoring guide and resolving grade discrepancies, while imposing demands and constraints, allows the enjoyment of highly specific pedagogical discussions. This meets a need of professional development expressed elsewhere, for example in the Gregoire Report.

Benefits to Students

It was not, however, only the teachers who benefitted from the process of holistic evaluation. Learning was enhanced as much as teaching was.

Particularly when a clarification of objectives leads to a revision of the test question, students have a clearer sense of what they are being asked to do and may therefore write more meaningful responses. Following one of the hypotheses of our survey, this perception of a writing assignment as a meaningful task---this involvement in the act of making meaning---leads students to make fewer grammatical errors.

Students' use of language is further strengthened by making the criteria explicit, allowing them to be partners in the testing process. They respond postively when their teachers share the scoring guide with them in advance. More specifically, they are encouraged to plan better once they are more aware of the expectations of the test. This planning helps them focus on the topic and finally deepens their understanding of the subject. Overall, students are made to feel part of an interpretive community and this in itself is a means of empowering them to do

meaningful writing.

A further means of improving learning is added when students are at least partially responsible for assigning grades. Their participation not only strengthens class dynamics by showing that grading is done fairly but also helps them become more active learners by having them apply the criteria. In addition, their understanding of the topic can be improved by seeing a range of possible interpretations at the same time as reviewing the material that has been tested. More specifically, reading other students' work helps them analyse the question posed.

In the long term learning is enhanced by the students' coming to see grades as being less arbitrary, more rational. They respond positively to the objectivity and anonymity of this system. Where comments are given as well as number grades, they appreciate these being positive rather than negative and the fact that they receive two or three rather than one. Students frequently assume they must figure out what the teacher wants rather than what the topic is; the opportunity to see that explicit criteria are used in grading an assignment is another important advantage of holistic evaluation. This forces them to analyze the question or define the task rather than to personalize assignments.

CONSTRAINTS

In a context other than that of mass testing, the process of holistic evaluation as generally described, for example by White, must be considerably modified.

First and foremost, there is no table leader or final adjudicator of grading discrepancies. This absence obviously undermines reliablity. However, reliability is not a concern in a classroom situation. Indeed, one strong reason for applying a proven process such as holistic evaluation to the classroom is to return responsibility for testing to teachers. If teachers can resume this responsibility, there will be less need for large scale testing and the standardization, including reliability, that it implies.

The lack of an adjudicator gives the scoring guide a central role in the process. This in turn limits the implementation of holistic grading in two ways. First, it precludes certain types of questions. Open-ended or "creative" questions do not allow the degree of specificity required in an effective rubric. On the other hand, questions that merely ask for recall of facts may lead to a check list which denies the holistic concept. And certain very personal issues may not be appropriate for reading by more than one marker. Second, the rubric forces a decision on the relative value of divergent abilities, such as compilation of data versus analysis. Teachers must be willing to examine and specify exactly which abilities they wish to test.

Timing is a second constraint. Holistic evaluation is inappropriate early in the term for two reasons. First, before the teacher has seen what the students in this particular course are capable of, s/he may draw up a scoring guide based on unrealistic expectations. Second, there is no personalized response or diagnosis of the students' learning. A mid-term test returned with the scoring guide as a model answer and an incentive to improvement seems the earliest opportunity to use the process. However, an end-of-semester test when students are more concerned with numbers than feed-back would be the most appropriate timing.

A third constraint is course structure. Particularly in some large, multi-sectioned courses there is no room for an extra assignment. In other cases, the department course description allows no variation of assignments between sections.

Finally, holistic evaluation is demanding. Collaborating on formulating the question and drawing up the rubric, then exchanging the essays that have been written demands time and organization and imposes deadlines. The procedure requires early preparation, which may negatively lead one to teach to the test, and the logistics of this marking system may even delay the return of the written tests to the students. Given some unforeseen circumstance such as an illness, these constraints can lead to frustration if not resentment. These problems are not, we found, balanced by the holistic reading of papers itself being faster than conventional modes of evaluation. And if the process is subverted, for example by not designing a scoring guide, the outcomes---probably a threatening number of discrepancies---can be even more negative. All these problems are aggravated when the department's prevailing ethic is one of individualism.

MODELS

Teachers must, then, be aware of the demands and limitations of the process in order to be able to benefit from its advantages. This balancing of the limitations against the benefits may lead to intermittent rather than regular use of the process. Several of our participants saw this as an excellent training---or re-training---exercise rather than an on-going responsibility. Whether used regularly or not, the holistic evaluation procedure can definitely enable a collaborative, organic teaching of "making meaning" which will in turn foster a "literacy of thoughtfulness" (Brown).

The following models are applicable to both in-class testing and take-home situations. Also, the assignments may be written individually or by groups. However, because of the constraints of the scoring guide, they may be more difficult to apply to long research papers. While they may be used at other times, these models are most applicable to end-of-term assignments which do not require a large amount of feedback to the students. For example an essay given earlier in the semester might be used for

diagnostic purposes, especially if the scoring guide is discussed with the students.

1. Two or More Teachers

The instructors do not need to be teaching the same course or even be members of the same department. They must, however, be both willing to work collaboratively and truly familiar with the material the other(s) is (are) teaching.

At the beginning, the teachers decide on a clear time frame for the preparation, writing, and grading of the test; this must fit their teaching and their personal schedules. Then they work together to design the written assignment and draw up a rubric. It is useful to design the scoring guide before distributing the assignment to students because the process of writing it may suggest revisions to the assignment itself. If the scoring guide is generalized, as in Appendix 2A, it may (as in Model 2) be discussed with the class. If it is specific, as in Appendix 2B, it can also be discussed if specific content demands are omitted. The teachers also have to agree on a "conversion" table that translates holistic grades into percentages. Appendix 7 shows prototypes of these. At a second meeting, they revise the scoring guide and then the question itself in order to ensure that they are clearly expressed and fulfil the purpose of the test objectives.

Once the essays have been written anonymously (identified only by student number), the two (or more) teachers together read three or four sample or "anchor" papers in order both to validate the scoring guide in terms of these particular students and to internalize its criteria for themselves. If the teachers can spend two or three hours together to read a full set, a controlled reading can then be established . In this case, they will read sets (say of 6) and then exchange grades in order continually to check the objectivity of their reading as well as to resolve discrepancies. This investment of time is subsequently rewarded by a noticeably faster reading of the entire set of papers. Some teachers, particularly once they are more comfortable with the process, may find marking more efficient and scoring more reliable without these frequent breaks and may prefer to wait until all the papers have been marked before resolving discrepancies. This will also be the case if the teachers prefer to read the papers alone.

Whether in a joint one-day or in a separate reading, papers are read first by the person who does not teach that class. This sequence is preferable logistically and philosophically. It allows the writer's teacher to add personal comments, if desired, without prejudicing the other reader; and, given that we often mark our own class more severely than another, it also provides a more objective assessment of the range. Grades are recorded either on a list of student numbers or on the back of the text. Comments by the first reader, if desired, are written on index

cards.

In the case of a separate reading, the essays are read on the second day by the writer's own teacher, who similarly records the grade but can put any desired comments on the text itself. The following day the teachers meet to discuss discrepancies of more than one mark. They seek to resolve these with reference to papers which they agreed had a score of, for example, 3 and 5; failing consensus, the decision of the student's teacher prevails.

At the end of either process, the papers are returned to the students accompanied by the converted grade and any comments made.

2. A Single Teacher

A teacher who is isolated either by being in a very small department or by giving an unusual course needs to substitute points of reference for the collaboration described above.

When designing the question and scoring guide, these points of reference can be provided by student responses to previous tests. By keeping copies of strong, mediocre, and weak student papers as well as notes on their assessments of these scripts, teachers can provide themselves with concrete criteria to use in revising the new test question and in drawing up the scoring guide. Where a generalized rubric such as Appendix 2a is used, it can be discussed with the class before the test. This not only helps the teacher revise the assignment but also clarifies the students' expectations of the test. Where (as in Appendix 2b) the rubric is content specific, the introductory section of the top scores can be discussed in order to show the criteria for superior answers.

The essays having been written anonymously (identified by student number rather than name), the teacher randomly selects ten to fifteen percent of the set as sample, or "anchor", papers. These are slowly read and compared to each other in order to validate the scoring guide previously designed and to internalize those criteria. After this, the full set of essays is read holistically, referring back to the validated scoring guide whenever there is a feeling of losing perspective. The holistic grades are recorded on a list of student numbers. Comments, if deemed appropriate, are written on an index card also identified by student number.

Before the next class, duplicates of each script are made for use by the students. To give students a sense of the grading range, three or four papers are read and discussed in terms of the scoring guide. Then two copies of each script are given to pairs of students, making sure that no writers read their own papers. If the teacher has two sections of the same course, objectivity can be greatly increased by the exchange being

between the two classes. In either case, each member of the pair independently reads and grades a paper; the pair then shares the grades given and discusses discrepancies in order to reach a consensus. The pair's agreed upon rade is then orally transmitted to the teacher, who records it on the class list. Where there is a discrepancy of more than one point between the pair's and the teacher's grade, the paper is read aloud and a consensus sought from the class as a whole. Failing consensus, the students' grade may raise, but not lower, the teacher's decision.

3. Two or More Teachers Plus Students

The closest facsimile to the mass testing procedure can be provided by a composite of Models 1 and 2.

The appropriate teachers work together (as in Model 1) to design the essay question, the scoring guide, and the conversion table. Each teacher ensures that the students write anonymously and afterwards makes duplicates of the papers to give to the other teacher for use in student evaluation.

The two teachers will first (as in Model 1) read a few "anchor" papers in order to verify criteria. This investment of time is subsequently rewarded by a noticeably faster reading of the entire set of papers. The first reader of the entire set of papers (as in Model 1) is the "other" teacher, but the second reader is students from the other class; they work in pairs to evaluate the papers (as in Model 2). The fact that the writer of the paper is not a classmate increases the objectivity of this assessment and undercuts the desire to give higher marks in the hopes that the person reading their paper will be equally generous. The pairs' grades are orally reported to the teacher and discrepancies discussed (as in Model 2). The list of holistic grades and the comments written on index cards, if these have been made, are then returned along with the scripts to the writers' teacher, who converts the grades before returning all data to the students.

CONCLUSION

Within a classroom context, these three models follow the six procedures specified by White (pages 404-405) as essential to a holistic evaluation process.

1. Controlled Essay Reading

While the readings are not as rigorously controlled in time and place as those of mass testing, our models do ensure that marking follows similar procedures. More importantly, these classroom readings allow the possiblity of creating a sense of an interpretive community condusive to learning rather than testing.

2. Scoring Criteria Guide

Given the lack of table leaders and of a communal time and place for reading in our models, heavy emphasis has been placed on the scoring guide as a means of ensuring objectivity and self-testing. While the classroom situation dictates that the guide must be constructed ahead of time rather than after reading the scripts themselves, our rubric still "reflects the reality of the actual test group" for the simple reason that it is prepared by teachers who know the particular students to be tested. In addition, it is validated in terms of these students' written products by the preliminary shared reading of "anchor" papers.

3. Sample or Anchor Papers

The purpose of these is to "calibrate" the reader. Our models force the teachers to take individual responsibility for this by internalizing the criteria from the scoring guide. Thus the shared reading of the "anchor" papers both validates the design of the rubric and calibrates the readers.

4. Checks on the Reading in Progress

While there is no outside adjudicator of readings, our models demand rational discussion of discrepancies between at least two individuals. Each must justify a grading decision in terms of the criteria previously agreed to.

5. Multiple Independent Scoring

Models B and C provide three readers, but A may have only two. However, the full discussion of discrepancies balances that lack of multiplicity. Teachers will be negotiating in order to ensure fair grading because they know these individuals. For while each script is anonymous, the class as a whole is personally known. Where this identification of a group is negative, the discussion of criteria with either the students or the other teacher will force such a bias to be exposed, hence overridden.

6. Evaluation and Record Keeping

Our models provide the possibility of a considerable amount of evaluation feed-back to the students. Also, particularly in Model B, teachers are encouraged to keep notes on their readings of the papers in order to have a record of their own grading reliability. All three include more extensive grading feedback to students that is normally the case in the classroom and all that is needed for a semester's records. Where latitudinal profiles are needed, such statistics are in no way precluded.

In summary, we have here a feasible means of adapting a proven procedure to the classroom. In so doing, teachers can foster the development of writing-to-learn across the curriculum

by helping students appreciate the "fundamentally social" (Witte, 1988) nature of reading and writing and see how the effectiveness of language is defined and determined by a community.

CHAPTER VI: DIFFUSION OF RESULTS

Locally we have already begun to make holistic evaluation and our study known. As reported in Chapter III, some of the preliminary work of the project involved publicity which allowed considerable discussion of this form of grading and of writing-to-learn in general. The circulation of the questionnaire for the faculty survey further increased sensitization to these issues. The project was also made known early in the year to the community as a whole by a two-page article in Communique, the faculty association's publication. Its progress was regularly reported to the Language Across the Curriculum Committee, whose minutes are circulated with the bi-weekly package for Academic Council meetings, and specific reference was made to the endeavour during an LAC presentation to the Board of Governors.

Specific plans have been made for 88-89. First, we wish to share our findings with our participants and the departments which have requested feedback. Copies of the report will be available to those who are interested, and we will visit any departments that invite us to highlight our findings. In addition, we hope to continue the collaborative aspect of the project and hold a closing meeting of the participants so that they can directly discuss the varieties of their experiences with holistic evaluation, particularly the types of questions they found appropriate and the criteria to include in scoring guides. Second, we wish to acknowledge our appreciation of the 158 people who so thoughtfully completed the survey questionnaire and will at the beginning of the semester be circulating a three-page summary of its main findings .

Continued discussion of one aspect of holistic evaluation---criteria---is guaranteed next year. Shared grading criteria, standards "across the board" in the words of one of our participants, have been adopted as an 88-89 priority by Language Across the Curriculum. This means that their meetings with departments and probably one of their Bulletins will focus on this topic. The subject has already been chosen as the focal point for the annual LAC workshop in January 1989 and may be proposed to Faculty Professional Development Committee as a key issue for Pedagogcial Day 1989.

Provincially a certain amount of liaison has begun. We have corresponded with a member of the Faculty of Education at Universite de Sherbrooke who is investigating the applicability of the concepts of Language Across the Curriculum to the French sector, and we visited the Centre d'aide en francais at Edouard Montpetit and discussed their language policy. We hope to maintain and expand these contacts by publishing a summary of our findings in Pedagogie colegiale and, if accepted, making a

presentation at one of the bufet-conferences of AQPC. To reach a different audience, we will also submit an article to Cegeppropos. In the English sector, we shall propose a workshop to "Springboards", the annual conference of the Association of Teachers of English in Quebec.

Outside the province we shall bring our findings on the clasroom applicability of holistic evaluation to the organization most concerned with the process, the National Testing Network in Writing, an American association which prides itself on an international membership. Our attendance at their conference this year not only deepened our understanding of the concept and informed us of the latest developments in the field but also helped shape the context for the presentation we shall propose for their 1989 meeting. We will also seek to make a similar presentation at "Inkshed", a Canadian conference concerned with writing.

CHAPTER VII: BUDGET

AMOUNT GRANTED			4000.00
EXPENSES			
Preparation of proposal		110.05	
Intervention			
Workshop 1	749.00		
Workshop 2	447.39		
Follow-up meeting	12.89		
Total		1209.28	
Liaison			
NTNW Conference	1201.78		
Phone calls	65.23		
Library loans	28.40		
AQPC membership	25.00		
Translation of summary	245.00		
Total		1565.41	
Printing			
Questionnaires	35.10		
Multiple purpose answers	21.40		
Publicity flyers	14.60		
Workshop packages	22.62		
Mailing labels	5.50		
Report (36 copies)	150.00		
Summary (110 copies)	50.00		
Total		299.22	
Diffusion of Results			
NTNW Presentation	240.00		
Inkshed Presentation	400.00		
Cegepropos translation	200.00		
Total		840.00	
TOTAL SPENT		4023.96	

APPENDIX 1: PUBLICITY FLYER

HOLISTIC EVALUATION: ONE MEANS OF IMPLEMENTING THE PRINCIPLES OF LANGUAGE ACROSS THE CURRICULUM

The assumption underlying and motivating this study is that, because it is a powerful means of learning, writing should be integrated throughout the curriculum. We use language to organize our perceptions of experience, to make these perceptions meaningful. So, when students write about history, for example, they are making meaning of---forcing themselves to understand---the concepts that have been introduced in their history course. In the words of one researcher, we can "view writing not only as a medium of thought but also as a vehicle for developing it. Some types of elaborated thinking are not possible unless the propositions are written down."

Many teachers share this assumption but do not give written assignments for the understandable reason that traditional grading methods are so time-consuming. Our project will provide participants with a solution to this problem by introducing them to an effective and efficient system of grading known as "holistic evaluation." This is a guided process in which a minimum of two readers respond immediately to a piece of writing and rank it, usually on a scale of 1 to 6, bearing in mind a set of criteria agreed upon earlier. Assignments are marked as a whole, without any subscores. The marker focuses on the overall strengths of the assignment, on the student's overall comprehension and ability to organize course material, rather than on individual errors or weaknesses.

There are several advantages to this system. For the teacher, it provides both a practical means of evaluating the student's overall grasp of course material and an opportunity to work closely with a colleague. For the student, there is the opportunity of active learning by developing the writing and thinking skills of a particular discipline.

Steps in the Process:

1. Preparation
   -Teams of two or three teachers of courses with shared objectives are formed.
   -Participants design a written assignment and define the criteria for grading.

2. Marking session
   -Papers are evaluated by two markers working independently.
   -Papers with a difference in scores of plus or minus one point are adjudicated.

SAMPLE QUESTION FOR HOLISTIC EVALUATION

The Question

Archaeologists have used the term "human" to refer to 20th century people as well as to creatures which lived 40,000, 100,000, 1,000,000 and even 2,000,000 years ago. Books on human evolution thus refer to Homo Sapiens Sapiens, Homo Sapiens Neanderthalensis, Homo Erectus and Homo Habilis. Since it is obvious that these creatures are not similar in either their physical or cultural adaptation, would it not be more accurate and more appropriate to limit the use of the term "human" to 20th century people?

Instructions

You will have 60 minutes to plan, write and revise an essay to answer this question. Atleast two evaluators will later read your essay and assign it a score. They will particularly be looking to see whether you show:

-understanding of the term "human"
-knowledge of the physical and cultural adaptations of the stages of human eveolution
-clear, well organized expression of these ideas

Some Suggestions

1. Take a few minutes to think about what you might say before you start to write.

2. On the left hand side of the examination booklet, write down a preliminary plan or proposed outline.

3. When you write your essay, leave a space between lines so that you will have room to make changes later.

4. Leave atleast 10 minutes at the end to read over your whole essay and make changes. You may cross out and add information as necessary, but make sure that your handwriting is legible.

5. In the last few minutes, read the essay word by word and correct spelling, punctuation, and grammar.

SCORING GUIDE FOR THE "HUMAN" ESSAY

PASS

6

The answer shows a superior understanding that:

1) the term "human" applies to all stages because the creatures all share basic characteristics
   -the similarities are fully described and shown to outweigh the differences
2) qualifiers are added to the term to make clear the physical and cultural adaptations at each stage
   -all adaptations for each stage are both described and assessed for their significance.

The essay is powerfully organized (analytically in terms of the concepts above or chronologically for the 4 stages) to express these perceptions, all of which are explicitly generalized in an introduction and conclusion.

5

The answer demonstrates an accurate grasp of the basic concepts:

-the similarities are described but not weighed against the differences

-all adaptations for each stage are described but not assessed for their significance.

The essay is clearly organized to express these perceptions which are generalized in an introduction and conclusion.

4

The answer demonstrates a commonplace understanding of the basic concepts:

-the similarities are all mentioned but not fully described

-all adaptations are listed but not described.

The essay is organized but contains errors that may cause the reader some distraction and even a shift in position from the introduction to the conclusion.

FAIL

3

The answer demonstrates only limited understanding or a partial misunderstanding of the basic concepts:

-only 2 or 3 similarities are listed or most are misinterpreted

-only 2 or 3 are listed or they are misinterpreted

The essay is organized in a way that impedes the reader's understanding, for example with some disagreements between the introduction and conclusion.

2

The answer demonstrates definite misunderstanding of the basic concepts or contains unsupported generalizations:

-2 or 3 inaccurate similarities are listed or no examples are given

-2 or 3 inaccurate examples are given or there are none.

The essay is organized so as to cause the reader significant difficulty.

1

The answer shows virtually no understanding of the concepts or a misreading of the question posed.

APPENDIX 2: SAMPLE SCORING GUIDES

I. GENERAL

Pass

8. A nearly perfect answer - fully enlightened showing a complete understanding of the question; proper usage of key artists and works of art. The answer has to be logically argued, the writing must be superb, and the examples must be well chosen to support the thesis. Reading and lecture notes must be intelligently keyed into the argument.

7. A solid answer - but not necessarily as superior or creative as the "8" answer. This answer may be excellent in some areas (as content, factual knowledge) but not as closely or carefully argued - but it must be well written and carefully thought through. An answer that could also not have beeen fully completed - but excellent as far as it went.

6. This is the answer that has something going for it that lifts it above just a median response - one or two good elements - but with obvious omissions and not adequate supportive material. This will be less well written than a "7" answer.

5. An answer that doesn't deal adequately with the basic question and in addition is not complete and fails to use intelligent examples. This answer will be minimal and factual - with little or no continuity or development - although it will show some knowledge of the question.

Fail

4. An answer that shows no real insight and a very limited understanding of the nature of the question. This will be basically a superficial answer and will exhibit some writing problems as well as being an incomplete answer.

3. An answer that is substantial with supportive material but without any comprehension, or a misinterpretation, of the question - written poorly.

2. Just slightly better than "1" - very insignificant handling, little or no content, and poor writing. No real understanding of the question.

1. An answer that has literally nothing to do with the question and where no knowledge is evident.

II. SPECIFIC

Pass

8. An answer in this category will show both a superior understanding of the implications of the question and skilled use of the text in supporting the answer.

This is it: the answer that is right on the money. The 8 answer is essentially just what you were looking for. It may not be perfect, but it is indeed very close.

In addition, the answer will be distinguished by lucid and orderly thinking and may even introduce an original interpretation of the question. It will be virtually free from mechanical errors and will exhibit a sense of style.

7. An answer in this category will show a competent grasp of the question and will make good use of textual material in giving the answer.

It will correctly identify the central character of each epic and the culture he represents. It will present the main strengths and weaknesses of each of those characters (or the strengths and weaknesses of the cultures they represent). And finally, it will provide adequate supporting examples from the two text.

The answer may not be as thoughtful or as carefully reasoned as the 8 paper but will demonstrate a solid understanding of the question. Although the answer may be less carefully organized than the 8 answer, it will not be characterized by mere statement of ideas at a high level of generality. It will contain evidence of the writer's ability to organize information into unified and coherent units. It will be largely free from serious errors in mechanics.

6. An answer in this category will show an acceptable but commonplace grasp of the question and a merely adequate ability to use the text in giving the answer.

It will correctly identify the central characters and the cultures they represent, and will present their main strengths and weaknesses, but will be characterized by the high level of generality with which it explains the answers. The answer will not be as complete in its presentation of the characters' strengths and wekanesses, and in using the text the writer will present only the most obvious specifics.

The answer will be well enough organized for the reader to follow the argument, and in general the mechanics will cause only minor distractions.

5. An answer in this category will be **adequate**, no more. It will show some familiarity with the material but no real insight. The understanding of the material will be limited or may involve partial misunderstanding.

It will correctly identify the central characters but may fail to identify the cultures they represent. It may omit important factors or deal with irrelevant factors in presenting the strengths and weaknesses of the characters. Its use of textual support will be characterized by a general lack of specificity.

The organization of the material may sometimes be unclear. There may be substantial problems at the sentence level, though not to the degree that they seriously interfere with readability.

<u>Fail</u>

4. An answer in this category will be **incomplete**.

It may be that the answer deals quite well with one of the characters and works but fails to discuss the other. In this case, though the answer may be quite well organized and presented, it simply does not deal with the whole question.

3. An answer in this category will show a **definite misunderstanding** of the intent of the question. The question may be misconstrued, mishandled, or redefined to accomodate what the writer wants to say or is able to say.

In terms of content, it will become clear that the writer has misunderstood or ignored the question. The answer consists of whatever the writer knows about the character or literary work, whether it is relevant to the text or not.

There may be serious or frequent mechanical errors, giving the impression of distinctly inferior writing.

2. An answer in this category will be **skeletal**.

It may be that the answer shows familiarity with both characters, but the answer will omit the most important factors in dealing with their strengths and weaknesses, and will fail to provide textual support.

1. This category is reserved for the answer that shows **virtually no understanding** of the question and **no familiarity** with the texts involved. It is clear that the writer simply does not know the material.

APPENDIX 3: FACULTY SURVEY

A group from Language Across the Curriculum has received DGEC support for an action-research project to investigate the effectiveness and desirability of using "holistic evaluation" for mid-term and final assignments. "Holistic Evaluation" is a guided process in which at least two readers respond immediately, thus quickly, to a piece of writing and rank it according to criteria previously agreed to.

The first stage in this study is to describe the "atmosphere" in which the project began. In order to describe the college as a whole rather than just our own impressions, we need to find out what our faculty think about student writing, evaluation schemes and collaborating with colleagues. The team (Kerr Canning, Dan Gryte, Jacques Lavoie, Diane McGee, and Christine Starnes) hopes you will be willing to take five minutes to provide such information by replying to this survey. We hope to receive your answer by November 2. Your anonymous answers will add to a composite picture which aims to describe, not to evaluate.

In the three sections below certain ideas about writing, evaluating and collaborating are expressed. These ideas do not necessarily represent those of the research team. By circling A, U, or D, you are asked to signal the ones with which you agree (A) or disagree (D), or about which you feel undecided (U). When responding, please consider the students currently in your classes.

WRITING

1. Only students who will use writing in their future careers need to develop writing ability.

2. If students are taught what one does when one writes, they can improve their writing.

3. Grammar and spelling are the biggest problems in student writing.

4. Students deepen their understanding of a topic when they write about it.

5. The main purpose of having students write is to let them show the teacher what they know.

6. Teachers outside English departments do not have time to help students with their writing.

7. Students can't write because they have had little practice in doing so.

8. Learning more grammar does not necessarily improve a student's writing.

9. After writing, one knows more about a topic than when one began.

10. I would like to assign more writing in my class, but I don't have time to grade it.

EVALUATING

11. Multiple-choice questions test understanding of the subject.

12. All the errors in student writing should be corrected by the teacher.

13. The "English" and the content of written assignments should be graded separately.

14. Comments about strengths foster improvement as much as pointing out errors.

15. It is good to use the same criteria for grading every written assignment.

16. Students should learn to make connections between the separate facts they learn in a course.

17. Written assignments are a valid means of evaluating students' learning.

18. It is more important to test recall of facts than the ability to see relationships between those facts.

19. Students study more before an objective test than before an essay test.

20.I feel confident assessing how well a student has fulfilled the objectives of my course.

COLLABORATING

21. Working in pairs or groups might undermine each teacher's autonomy in the classroom.

22. Interaction between members of my department usually deals with pedagogical matters.

23. I enjoy discussing teaching methods.

24. Working in my department, I feel I am a member of a team with common goals.

25. Regulations in the Cahiers limit my ability to introduce innovative teaching.

26. Exchanges about teaching methods are usually not productive.

27. My department does much to foster co-operation of teachers in the development of pedagogy.

28. At John Abbott, there should be more opportunities for exchanges on teaching methods.

APPENDIX 4: STUDENT QUESTIONNAIRE

On the general purpose answer sheet, please respond to each of these statements. Fill in A if you agree, B if you are undecided, and C if you disagree. None of these statements is right or wrong.

I. WRITING

1. Written (essay) assignments let me show the teacher what I have learned.

2. I would be able to write better essays if did it more often.

3. Grammar and spelling errors are the worst weaknesses in writing.

4. I learn about a topic when I write about it.

5. Organising an essay clearly is more important than writing "good" sentences.

6. When I was shown what to do to write a paper, my writing improved.

7. Writing is the process of putting on paper what is stored in your mind.

8. Teachers outside English departments do not care about the quality of your writing.

9. I do not think my writng will improve in college.

10. I feel satisfied when I have finished a piece of writing.

EVALUATION

11. I study more before a multiple-choice or short answer test than before an essay test.

12. At the end of a multiple choice or short answer test I feel I haven't made use of all I know.

13. I like it when a teacher gives separate marks for the content of a paper and the quality of the writing.

14. It is helpful when teachers point out what I have done well.

15. I rarely go over the corrections a teacher makes on my paper.

16. Teachers tend to evaluate essays on the basis of their opinion of the students, not of the paper.

18. I have to remember more in order to write an essay test or examination than to write a multiple choice test.

19. It is unfair for grammar, spelling and punctuation to count for marks except on English assignments.

20. When I put my ideas into an essay, I am showing what I have learned.

THANK YOU FOR YOUR TIME.
PLEASE ADD YOUR PERSONAL COMMENTS BELOW.

APPENDIX 5: INTERVIEW QUESTIONS

CLOSED

1. Learning Objectives

a. Did the question test the learning you had predicted?

b. Did you feel comfortable with the grades that resulted from the test?

2. Teaching Objectives

a. Did the students take the question seriously?

b. Did the students complain about the mark?

3. Teaching Process

a. Did "holistic evaluation" save you time?

b. Did you find setting up the scoring guide useful:
to prepare for the grading?
to revise the question?

c. Did you find answers that did not fit the scoring guide?

4. Collaboration

a. Did you work with someone you do not normally collaborate with?

b. Did you feel comfortable discussing these questions so specifically?

c. Did you have many disagreements (more than one point) on the grades?

d. What do you think caused these discrepancies?

e. Were you comfortable dealing with these differences?

5. Summary

a. Would you do it again or recommend "holistic evaluation" to others?

OPEN

1. In general, how do you feel about the process?

2. What do you feel you students learned from it?

3. What do you feel you learned from it?

APPENDIX 6: SHARED GRADING CRITERIA

"GOOD" WRITING

It is obvious that subject expertise is at the base of our expectations of student writing. Beyond this, four factors inform our decisions to move a student's grade up or down.

1. Reasoning

There is a basic need for a clear direction of argumentation and an adequate and appropriate use of evidence. This need does not change with different writing tasks.

2. Organization

The importance of organization stems from the reader's need to have concrete evidence of the writer's reasoning.

3. Appropriateness

Whether a student is meeting this criterion is most clearly shown by the register or level of formality of the words chosen but also depends on syntax and structure.

4. Mechanics

Where these are inaccurate, there is a communication breakdown, so the paper cannot fulfil its purpose.

These four factors have to be weighted according to the general learning objectives of the course and the specific objectives of the test or assignment. Importantly, we must guard against well-written emptiness scoring well.

APPENDIX 7: CONVERSION TABLES

Below are three suggested tables. The choice between these or of another must be determined by the normal distribution of grades for the particular test in the given course. The findings from Chemistry suggest that these will vary significantly between unit tests and final examinations.

	Holistic Score	Percentage
Table 1		
PASS	5	100
	4	80
	3	70
	2	60
FAIL	1	50
Table 2		
PASS	6	92
	5	80
	4	68
FAIL	3	56
	2	44
	1	32
Table 3		
PASS	6	90 (96)
	5	80 (84)
	4	70 (72)
		60
FAIL	2	50
	1	40

ERIC Clearinghouse for Junior Colleges
JUN 21 1990